

**Міністерство культури України**  
**Харківська державна академія культури**

Кафедра інформаційних технологій

**“ЗАТВЕРДЖУЮ”**

Проректор з навчальної роботи

“ \_\_\_\_\_ ” \_\_\_\_\_ 20\_\_ року

**Прикладна лінгвістика**

(назва навчальної дисципліни)

**Конспект лекцій навчальної дисципліни**

Освітня програма

Інформаційна та документаційна діяльність  
першого (бакалаврського) рівня освіти

спеціальність 029 Інформаційна, бібліотечна та архівна справа

спеціалізація Інформаційна та документаційна діяльність

факультет Соціальних комунікацій

Харків, 2018

Конспект лекцій Прикладна лінгвістика для студентів-бакалаврів, спеціальність 029 Інформаційна, бібліотечна та архівна справа.

„05” грудня, 2018 року 63 с.

Мова навчання – українська

Розробник: Брусенцев В.О., доцент кафедри інформаційних технологій, кандидат технічних наук, доцент

Конспект лекцій затверджений на засіданні кафедри інформаційних технологій

Протокол від “05” грудня 2018 року № 7

Завідувач кафедри проф., д.т.н. Асєєв Г.Г.

\_\_\_\_\_ ( \_\_\_\_\_ )  
(підпис) (прізвище та ініціали)

“ \_\_\_\_\_ ” \_\_\_\_\_ 20\_\_ року

Схвалено Методичною радою факультету соціальних комунікацій за спеціальністю 029 Інформаційна, бібліотечна та архівна справа

Протокол від “10” грудня 2018 року № 5

Голова Методичної ради \_\_\_\_\_ (Перевозна Т.О.)  
(підпис) (прізвище та ініціали)

## 1. План лекційних занять

№ теми	Тема лекції	Кількість годин	
		ден.	заоч.
<b>Розділ 1. Теоретичні особливості прикладної лінгвістики</b>			
1	Лінгвістика як наука: теоретична та прикладна	2	0,5
2	Семіотика як засіб комп'ютерної лінгвістики	2	0,5
3	Штучні мови. Формальна граматики	2	1
4	Лінгвістична модель та моделі подання знань	2	1
Разом за розділом 1		8	3
<b>Розділ 2. Практичний аспект прикладної (комп'ютерної) лінгвістики</b>			
5	Інформаційний пошук як засіб прикладної лінгвістики	2	0,5
6	Лексикографія та словники	2	0,5
7	Машинний переклад та технології	2	1
8	Автоматична обробка тексту. Лінгводідактика	2	1
Разом за розділом 2		8	3
<b>Усього годин</b>		16	6

## РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСОБЛИВОСТІ ПРИКЛАДНОЇ ЛІНГВІСТИКИ

### Лекція № 1

Тема 1. Лінгвістика як наука: теоретична та прикладна – 2 год.

Мета лекції: визначити загальне місце курсу. Провести аналіз завдань та напрямків прикладної лінгвістики.

#### План лекції

1. *Лінгвістика як наука. Складна система лінгвістичних дисциплін та напрямків.*
2. *Теоретична та прикладна (практична) лінгвістика. Завдання та напрямки прикладної лінгвістики.*
3. *Зв'язок лінгвістики з іншими науками – природними та гуманітарними.*

#### Основний зміст

1. *Лінгвістика як наука. Складна система лінгвістичних дисциплін та напрямків.*

Великий енциклопедичний словник «Мовознавство» дає наступне визначення лінгвістичній науці: «Лінгвістика (мовознавство) – це наука про природну людську мову взагалі і про всі мови світу як індивідуальних його представників». Лінгвістика в широкому сенсі займається пізнанням мови й передачею результатів цього пізнання іншим людям для їхніх практичних цілей.

Цікавий факт, що сам термін «лінгвістика» (linguistics) увійшов у науковий побут в 1847 році, хоча термін «linguist» (у значенні «a student of language») виник на 200 років раніше.

Мовознавство виникло у зв'язку із практичними потребами людей і тісно пов'язане з появою писемності в період, за різними

оцінками, від 2 до 5 тисяч років до нашої ери на Древньому Сході. Одними з перших лінгвістичних продуктів вчені вважають шумерські глосси (25 ст. до н.е.), а першою серйозною науковою працею – формальні граматики санскриту давньоіндійського вченого Паніні (V ст. до н.е.). Саме ж мовознавство стало оформлятися як окрема наука лише в XVII-XVIII століттях, що було обумовлено її величезною складністю і недоліком знань про мову – об'єкті цієї науки. В XIX-XX ст. відзначається підйом в області мовознавства, з'являються відомі лінгвістичні школи (американська, празька, копенгагенська, московська та ін.) і напрямки. Величезний внесок у лінгвістичну теорію вносить Фердинанд Де Соссюр, що розробляє основи структуралізму, або структурну лінгвістику, розвиває теорію мови і мовлення. Практичний напрямок, що має назву сьогодні «прикладна лінгвістика», активно став розвиватися в середині XX століття у зв'язку з появою комп'ютерних технологій та удосконаленням технічних засобів.

Мовознавство вивчає різні сторони мови, розглядаючи їх у рамках своїх окремих наукових напрямків. Наприклад:

Фонетика і фонологія – наукові напрямки мовознавства, що вивчають звукову систему мови та особливості її функціонування.

Лексикологія вивчає словниковий склад мови і закономірності його розвитку.

Грамматика, як наука про способи і засоби побудови та зміни слова, про способи і засоби побудови речення, підрозділяється на дві області – морфологію і синтаксис. Морфологія – навчання про способи і засоби побудови та зміни слів. Синтаксис – навчання про способи і засоби побудови речень.

Семантика – наука про зміст (значення) слів і речень, частин мови та членів речень.

Прагматика – навчання про умови і цілі комунікації, що впливають на розуміння, так називане «вивчення мови в контексті», вивчення відносин між засобами мови і тими, хто цими засобами користується.

Дискурс вивчає та аналізує характерні для даного виду текстів (або комунікантів) особливості лексики, синтаксису, семантики або прагматики мовних одиниць, що проявляються в актуальних

комунікативних актах, мовленні і письмових текстах (наприклад, спортивний дискурс).

Такий підхід традиційний у вивченні мови і нам частково він знайомий зі шкільного курсу викладання української та іноземної мов. Поглиблено такий підхід до структури мови та теорії мовознавства дається і у ВНЗ в рамках освітніх програм по лінгвістиці. Але ми також будемо розглядати лінгвістику як складну та розгалужену систему дисциплін як теоретичного, так і прикладного характеру. Наведемо лише деякі з таких дисциплін для демонстрації розмаїтості мовознавчих напрямків.

Приклади лінгвістичних дисциплін:

1. Загальне мовознавство займається проблемами, що стосуються всіх мов, досліджує сутність і природу мови взагалі, проблему його походження і загальні закони його функціонування, структуру, класифікацію мов. Загальне мовознавство розробляє методи дослідження мов, формулює мовні універсалії, тобто положення, дійсні для всіх або більшості мов світу.

2. Приватне мовознавство вивчає одну або групу родинних мов (наприклад, германістика, романістика, тюркологія, балканістика і т.ін.). Залежно від цілей і завдань дослідження приватне мовознавство може бути синхронічним, що описує фактор мови в який-небудь момент історії, в одній тимчасовій площині, найчастіше сучасний його стан, і діахронічним, що простежує історичний розвиток мови.

3. Порівняльно-історичне мовознавство (компаративістика) займається встановленням ступеня споріднення між мовами (побудовою генеалогічної класифікації мов), реконструкцією перших мов, або прамов, дослідженням різних процесів в історії мов, їхніх груп і сімей, етимологією слів.

4. Типологія (універсалізм) займається з'ясуванням найбільш загальних закономірностей різних мов, не зв'язаних між собою загальним походженням або взаємовпливом. У випадку, якщо деяке явище виявляється в представницькій групі мов, воно може вважатися типологічною закономірністю, застосованої до мови як такої. Типологічний аналіз можливий на рівні звуку (фонетична типологія), на рівні слова (морфологічна типологія), речення (синтаксична типологія) і далі (типологія тексту або дискурсу).

5. Діалектологія (лінгвістична географія) – вивчає місцеві територіальні різновиди однієї мови.

6. Когнітивна лінгвістика – міждисциплінарний напрямок мовознавства, тісно пов'язане із семантикою, що досліджує проблеми співвідношення мови, свідомості і мислення, роль мови в пізнанні, розумінні та відображенні навколишньої дійсності.

7. Паралінгвістика – вивчає немовні (невербальні) засоби в мовленні, що передають разом з вербальними змістовну інформацію в складі мовного повідомлення, а також сукупність таких засобів. Вивчення невербального спілкування є важливим компонентом міжкультурної комунікації. Паралінгвістику також вважають розділом невербальної семіотики.

8. Психолінгвістика – пов'язана з вивченням мовлення, мовотворення (найчастіше конкретного індивіду) для рішення тих проблем, які не можуть бути вирішені ні в психології, ні в лінгвістиці. Психолінгвістика – самостійний науковий напрямок, що виник на стику лінгвістики і психології. Об'єднання цих двох дисциплін в одну приміжну дисципліну дозволяє використати понятійний апарат лінгвістики для опису мовної форми мовленнєвих висловлювань і понятійний апарат психології для опису та пояснення психічних процесів виробництва і сприйняття мовлення. У рамках психолінгвістики вивчається, наприклад, дитяче мовлення, мовні відхилення при психічних захворюваннях.

9. Соціолінгвістика – вивчає роль мови в суспільстві, вплив суспільства на мову, мову у зв'язку із соціальними умовами його існування, комплексом зовнішніх обставин, у яких реально функціонує та розвивається мова: суспільство людей, що використовують дану мову, соціальна структура цього суспільства, розходження між носіями мови у віці, соціальному статусі, рівні культури та утворення, місці проживання, а також розходження в їхньому мовному поведженні залежно від ситуації спілкування. У рамках соціолінгвістики активно розвивається, наприклад, політична лінгвістика.

10. Прикладна лінгвістика – до цієї області ставляться практичні напрямки мовознавства, часто міждисциплінарного характеру, наприклад, комп'ютерна лінгвістика (машинний переклад, лінгвістичні основи інформатики, інформаційний пошук та ін.), лінгводидактика (навчання мові), дешифрування

(дослідження текстів на незнайомому коді, або мові, для одержання інформації), математична лінгвістика (наприклад, квантитативна лінгвістика, що проводить статистичний аналіз мови і текстів, що займається побудовою та використанням структурно-ймовірнісних моделей мов).

*2. Теоретична та прикладна (практична) лінгвістика. Завдання та напрями прикладної лінгвістики.*

Лінгвістику умовно прийнято підрозділяти на теоретичну, іноді її ще називають «наукова лінгвістика», або «теорія мовознавства» (у рамках цього напрямку розглядаються різні наукові концепції, лінгвістичні теорії, лінгвістичні школи, мова з погляду його структури та систем), і прикладну (практичну) лінгвістику.

Одним з найважливіших напрямків розвитку лінгвістики є прикладна лінгвістика, одним з головних завдань якого є рішення та оптимізація традиційних лінгвістичних завдань, як, наприклад, переклад або навчання мовам. Забезпеченню практичних потреб людей служать такі розділи прикладної лінгвістики, як написання практичних граматик, практична фонетика, практична лексикографія (створення навчальних або галузевих словників).

Кожен з напрямків практичної лінгвістики звичайно має своє відображення в сфері теоретичної лінгвістики. Так, наприклад, проблеми перекладу вивчаються в рамках такої дисципліни як перекладознавство (Translation Studies), що формує свою теорію перекладу, різні перекладацькі концепції.

У широкій інтерпретації прикладна лінгвістика є комплексною науковою дисципліною, що застосовує лінгвістичні знання в різних ситуаціях для рішення різного роду практичних завдань (таких, наприклад, як машинний переклад, технічна комунікація, розпізнавання і синтез мовлення, інформаційний пошук та ін.).

Слід також зазначити, що за кордоном термін «Applied Linguistics» іноді трактується вузько, а під ним часто розуміється тільки сфера викладання іноземних мов, тобто сфера підготовки викладачів іноземних мов, вивчення методики лінгводидактики.

На думку багатьох вчених теперішній період розвитку мовознавства характеризується серйозним впливом на лінгвістичну теорію прикладної лінгвістики. Також, коли говорять про перевірку



лінгвістичних теорій на практиці, то мають на увазі не тільки, наприклад, створення практичних граматики і словників, але й їхній ефективний додаток (використання) у навчанні та викладанні.

Сучасна прикладна лінгвістика активно розвивається й вирішує багато проблем обробки усного та письмового мовлення, вилучення та використання інформації з текстів, підвищує ефективність комунікації, у тому числі в сфері інформаційних і комп'ютерних технологій.

Прикладна лінгвістика являє собою ту сферу, де реально проводяться лінгвістичні експерименти, що мають метою верифікацію положень теорії мовознавства й перевірку ефективності лінгвістичних продуктів, що створюються розробниками.

Визначимо коло деяких основних завдань прикладної лінгвістики:

1. Переклад с/на іноземну мову.
2. Навчання іноземній мові (методики навчання), лінгводидактика.
3. Підтримка комунікації за допомогою технічних засобів.
4. Лінгвістичні основи інформатики (створення штучних мов, наприклад, мов програмування, web-розробки).
5. Лінгвістичні основи інформаційного пошуку.
6. Анотування, реферування, класифікація текстів.
7. Складання словників (практична лексикографія).
8. Термінологія і термінографія (упорядкування, стандартизація та уніфікація науково-технічної термінології).
9. Розпізнавання символів тексту.
10. Розпізнавання і синтез усного мовлення.

В останні 50 років впровадження нових інформаційних технологій в усі сфери людського спілкування прикладна лінгвістика розвивається за напрямком автоматизації основних завдань, оптимізації комунікації, тобто розвивається такий напрямок прикладної лінгвістики, що назвали комп'ютерною лінгвістикою. Сюди входять, наприклад:

- машинний переклад (Machine Translation);
- комп'ютерна лексикографія (підготовка електронних словників);

- комп'ютерна лінгводидактика (CALL/T – Computer Assisted Language Learning and Teaching);
- автоматична обробка природних мов (NLP – Natural Language Processing).

У рамках прикладної лінгвістики для рішення вищевказаних завдань необхідно було виробити ефективні методи дослідження мови. Серед таких класичних методів можна відзначити, наприклад, імовірно-статистичні методи, логіко-математичні методи, методи моделювання, корпусного аналізу, контент-аналізу та багато інших, які активно і плідно використовувались в інших науках і знайшли свій ефективний додаток у лінгвістиці.

### *3. Зв'язок лінгвістики з іншими науками – природними та гуманітарними.*

Лінгвістика – складна багатопланова область знань і додатків, що стикається як з багатьма гуманітарними науками, так і з природно-науковими. В останні роки з'явилися суміжні лінгвістичні напрямки на стику тих або інших наук, які активно розвиваються за рахунок нових теоретичних і практичних додатків. Вкажемо лише деякі:

- Соціологія, Політологія ⇒ Соціолінгвістика, Політична лінгвістика
- Філологія ⇒ Лінгвістика тексту
- Психологія ⇒ Психолінгвістика
- Логіка ⇒ Когнітивна лінгвістика, Інформаційний пошук, Моделювання семантики і знань
- Математика ⇒ Математична лінгвістика, Лінгвостатистика
- Біологія ⇒ Фонетика
- Фізика ⇒ Розпізнавання та синтез мовлення
- Історія, археологія етнографія ⇒ Етнолінгвістика, Дешифрування, Порівняльно-історичне мовознавство
- Юридична наука ⇒ Юридична лінгвістика розглядає мову права, проводить лінгвістичні експертизи текстів і висловлень)

- Інформатика ⇒ Комп'ютерна лінгвістика (наприклад, її розділ – лінгвістичні основи інформатики, пов'язаний з розробкою й обробкою штучних мов програмування);
- Семіотика ⇒ Знакова теорія мови, WEB-дизайн.

#### Запитання для самоперевірки

1. Встановити зв'язок між термінами „лінгвістика”, „граматика”, „морфологія”, „синтаксис”, „фонетика”, „фонологія”, „семантика”, „лексикологія”, „прагматика”, „дискурс”.
  2. Охарактеризувати основні завдання та напрямки прикладної лінгвістики.
  3. Визначити зв'язок лінгвістики з іншими науками.
- Література: [3, 4, 6, 7, 9, 11, 19, 26, 31].

### Лекція № 2

Тема 2. Семіотика як засіб комп'ютерної лінгвістики – 2 год.

Мета лекції: визначити роль семіотики. Провести аналіз завдань комп'ютерної лінгвістики.

#### План лекції

1. Семіотика. Знак і теорія знаків. Мова як мовна система.
2. Інформація. Інформатика.
3. Комп'ютерна лінгвістика.

#### Основний зміст

1. Семіотика. Знак і теорія знаків. Мова як мовна система.

Семіотика (греч. semeon – знак) – це наука про загальну теорію знаку, що досліджує будь-які знакові системи як засоби позначення і передачі значення або інформації.

Як окрема наука семіотика оформилася лише в ХХ столітті, але історія цього наукового напрямку досить цікава. Один з відомих сучасних учених і письменників Умберто Еко у своїй книзі

«Semiotics and the Philosophy of Language» відзначає, що теоретичні концепції про знаки явно проглядаються в роботах багатьох відомих світу мислителів і філософів, починаючи від Платона та Аристотеля.

Засновником семіотики, як формальної строгої науки, багато хто вважають американського логіка і філософа Чарльза Пірса, що жив в XIX столітті. Пірс визначив поняття знаку, розділив знаки на класи – індекси, ікони та символи – і встановив завдання і рамки нової науки. Ідеї Пірса більше 100 років були практично не відомі, а його теоретична праця була майже забута сучасниками, але допоміг випадок. Ім'я і праці Пірса одержали широку популярність в 30-х роках XX століття завдяки іншому американському вченому – Чарльзу Моррісу, що адаптував праці Пірса для широкого читача, фактично забезпечив їм і Пірсу всесвітню популярність, назвавши того основоположником семіотики.

Ч. Морріс у свою чергу покладав на семіотику більші надії. Він бачив у ній «знаряддя для об'єднання всіх наук». Оскільки людська культура заснована на знакових системах, вивчення таких систем є важливим завданням різних гуманітарних дисциплін (психологія, логіка, філософія, лінгвістика, культурологія та ін.). Однак Морріс розумів, що існує вагомий недолік теоретичної бази, що дозволила б узагальнити результати цих дисциплін. Цю проблему Морріс і намагався вирішити за допомогою семіотики як «науки про знаки».

Пірс, Морріс та їхні послідовники розвивали логічну лінію семіотики, а лінгвістична галузь семіотики, «вивчаюча роль знаків у житті суспільства», паралельно розроблялася в працях Фердінанда де Соссюра. Швейцарський лінгвіст Фердінанд де Соссюр також сформулював основи науки про знаки, назвавши її семіологією, і визначив її відношення до лінгвістики.

В 60 – 80-і роки XX століття внесок у розвиток семіотики внесли вчені з СРСР. До початку 1960-х років у Москві сформувалася група лінгвістів, що прийшли до семіотики різними шляхами: від структурної лінгвістики і машинного перекладу, компаративістики та загального мовознавства. Паралельно в м.Тарту в області семіотики працював всесвітньо відомий учений Юрій Лотман. Саме ці вчені стали ідеологами тієї семіотичної галузі, що згодом одержала назву Московсько-Тартуська семіотична школа.

Сучасна міжнародна асоціація вчених, що працюють в області семіотики, The International Association for Semiotic Studies, була утворена в 1969 році. Із цього часу і починається історія сучасної семіотики як складної багатопланової науки, що являє собою досить розвинену теорію, методи якої дозволяють аналізувати найрізноманітніші сфери людської діяльності. Сучасна семіотика має безліч відомих продуктивних напрямків розвитку, наприклад:

- семіотика культури;
- семіотика літератури;
- політична семіотика;
- семіотика масових комунікацій;
- педагогічна семіотика;
- семіотика тексту;
- комп'ютерна семіотика.

і інші цікаві напрямки та додатки.

Центральним поняттям семіотики є поняття знаку.

Визначити поняття «знак» намагалося багато вчених. Основоположник семіотики Ч.Пірс у своїх роботах дає 76 визначень знаку. Ми виберемо наступне визначення, що є найвідомішим визначенням знаку по Пірсу:

1. «Namely, a sign is something, A, which brings something, B, its interpretant sign determined or created by it, into the same sort of correspondence with something, C, its object, as that in which itself stands to C».

Це визначення в перекладі на українську одержало свою інтерпретацію – «Знак ... це щось, що стоїть для когось замість чогось іншого в тім або іншому плані або відношенні».

Також для первісного пояснення цього терміну серед численних визначень знаку в навчальній лінгвістичній літературі виберемо, що перегукуються з визначенням Пірса.

2. «Знак – це якесь А, навмисно поставлене кимсь замість деякого С з метою інформувати когось про це С».

Наприклад, значок копірайту © на книзі або сайті поставлено, щоб сказати про захист прав того, хто написав і видав цю книгу або розробив сайт.

Знак матеріальний і має спрямоване значення. У знаку виділяють два основних аспекти (або «плана»):

- план вираження (матеріальний, як виглядає);
- план змісту (що позначає, зміст).

Класичний поділ знаків за Пірсом включає три класи:

- 1) іконічні знаки,
- 2) знаки-індекси,
- 3) знаки-символи.

З погляду Пірса, іконічний знак – це знак, що володіє рядом властивостей, властивих позначуваному їм об'єкту. Відносини між знаком та об'єктом – це відносини подоби; знак виявляється знаком просто в силу того, що йому «трапилося бути схожим» на свій об'єкт (наприклад, креслення, схеми, діаграми, комп'ютерні іконки в меню редактора та ін.).

Знаки-індекси – вказівні знаки, у яких означуване та означаюче зв'язані між собою причинно-наслідковими відносинами. Пірс приводить такий цікавий приклад цього типу знаку: «Такий, наприклад, який-небудь предмет з дірою від кулі як знак пострілу; тому що без пострілу не було б ніякої діри, але сама діра там є, неважливо, чи дістане в кого-небудь розумових здатностей зв'язати її із пострілом чи ні».

Знаки-символи – знаки, у яких означуване та означаюче зв'язані між собою в рамках деякої конвенції, тобто за попередньою домовленістю. Національні мови, грошові системи, дорожні знаки – приклади таких конвенцій.

Юрій Лотман у свою чергу затверджував, що знаки діляться на умовні (символьні) і образотворчі (іконічні):

– умовний – знак, у якому зв'язок між виразом та змістом встановлено за згодою групи людей, «внутрішньо не мотивований». Найпоширеніший умовний знак в природній мові (ПМ) – це слово.

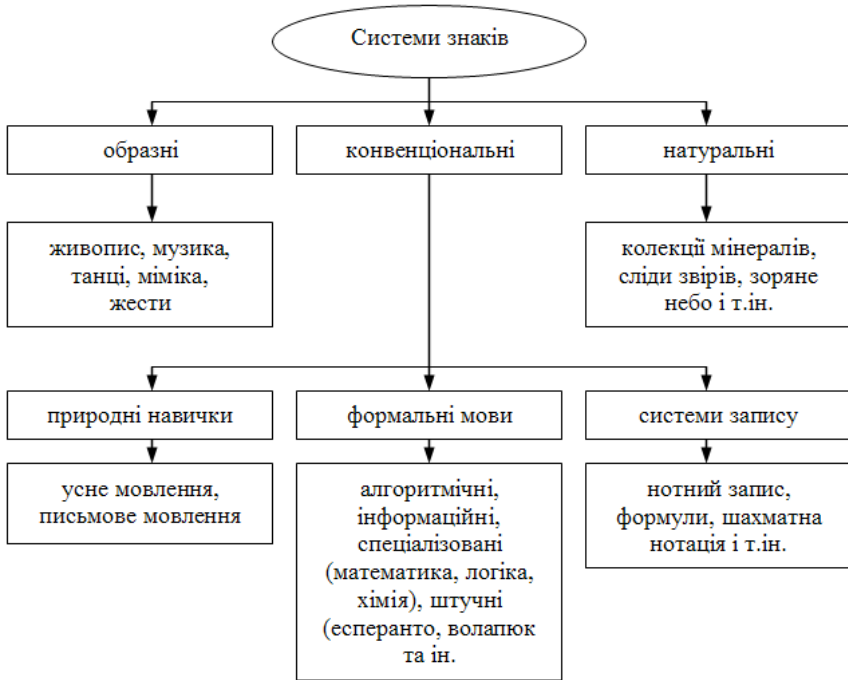
– образотворчий – знак, у якому значення має природно йому властивий вираз. Найпоширеніший образотворчий знак – малюнок.

Знак – звичайно це член певної знакової системи.

Наприклад, усім відомі такі знакові системи як дорожні знаки, сигнальні коди (абетка Морзе), креслярські позначення, комп'ютерні піктограми і т.ін.

Лінгвістику цікавлять загальні положення семіотики про знаки, розпізнавальні ознаки знаків, способи класифікації знаків, комбінація їх у систему для рішення конкретних завдань.

Визначаючи мову, часто говорять, що «Мова – це система знаків». Тут саме й проглядається чіткий зв'язок лінгвістики і семіотики.



Знаковими одиницями мови вважають морфеми і лексеми, тому що вони несуть інформацію та значення на відміну від одиниць нижчого рівня мови – звуків і букв (так називані знаки 1-го роду).

Лінгвістика також виділяє 3 аспекти знаку:

1. Синтактика (вивчає відносини між знаками).
2. Семантика (вивчає відносини між знаком і значимим).
3. Прагматика (вивчає відносини між знаками й тими, хто їх використовує).

Розуміючи, що мова – це деяка система знаків, і говорячи, що мова – це засіб комунікації (основне визначення в прикладній лінгвістиці), або засіб передачі значення (визначення семіотики), ми знаходимо ще один зв'язок лінгвістики і семіотики. Крім того, звідси виходить, що мова є засобом передачі інформації, тим самим

ми встановлюємо зв'язок з такою перспективною наукою, як інформатика.

## 2. Інформація. Інформатика.

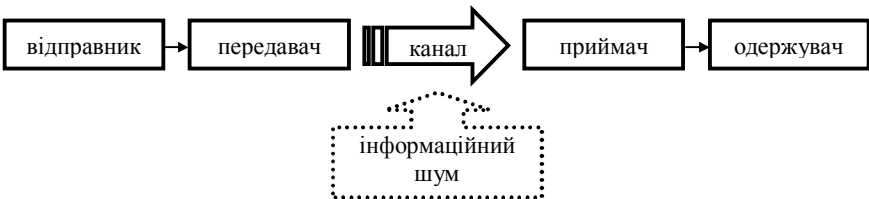
Спочатку дамо кілька визначень поняттю інформації.

– під інформацією розуміються не самі предмети та процеси, а їхні значення і характеристики (їхнього відображення у вигляді чисел, текстів, формул, таблиць, графіки), тобто говорять про текстову, графічну інформацію.

Цікаві і такі визначення:

- інформація є засобом усунення невизначеності в знаннях;
- Інформація – відомості, дані, що несуть значення для конкретної людини.

Питаннями інформації займається наука інформатика. Класична теорія інформації була обґрунтована К. Шенноном в 50-х рр. XX століття. Основною проблематикою теорії інформації є оптимізація процесів передачі інформації з каналів зв'язку від відправника до одержувача. Ці питання стали найбільш актуальні в наш час із розвитком усіляких засобів зв'язку, у т.ч. мобільних і комп'ютерних.



Інформація доступна:

- 1) сприйняттю (розпізнавання образів, текстів і т.ін.);
- 2) передачі (по каналах зв'язку) у вигляді сигналів (аналогових/цифрових);
- 3) обробці (перетворенню в зручну форму);
- 4) поданню (для демонстрації людині).

У середині XX століття у зв'язку зі зростаючими обсягами інформації та з появою нових інформаційних технологій і



середовищ склалась та стала активно розвиватися така наука як інформатика.

Під інформатикою (Computer Science) розуміють науку про закономірності запису, зберігання, переробки, передачі та використання інформації за допомогою технічних засобів.

Основні поняття в інформатиці – це код, система умовних знаків або символів; алфавіт – набір знаків коду, текст – послідовність знаків даного повідомлення. Термін «кодування інформації» – це подання повідомлення у формі, зручної для передачі даним каналом.

Більшість кодів для кодування інформації засновано на тих або інших системах числення. Передача і зберігання інформації при цьому – це робота із числами. При порівнянні систем числення та реалізації комп'ютерної техніки ефективною виявилася двійкова система, тому що логічні елементи повинні мати два стійких стани (включене/виключене або 1;0). З технічної точки зору, інформація – це сигнал (1,0) або елементарна подія «так» – «ні», вимірюване у двійкових одиницях – бітах (міра Хартлі).

Проте числа, хоча і зручні для передачі та зберігання в комп'ютері, не зручні для сприйняття і подання інформації для потреб людини. Універсальним же сховищем і засобом передачі інформації є природна людська мова. Крім того, що переважна частина інформації існує у вигляді письмових та усних текстів на природній мові.

### *3. Комп'ютерна лінгвістика.*

Сучасна людина живе в глобальному інформаційному середовищі, а обсяги інформації зростають на порядки щорічно. Сучасні інформаційні технології включають зростаюче число мережних комунікаційних технологій, автоматизованих інформаційних систем, засобів масової комунікації, систем інформаційного пошуку, систем машинного перекладу, а технічна комунікація розвивається у бік оптимізації інформаційних процесів і використання зручної для людини мови. Проблеми мовної комунікації «людина – комп'ютер – людина» і моделювання мови лежать в області досліджень такої молоді науки як комп'ютерна лінгвістика (КЛ, Computational Linguistics), що утворилася на стику

інформатики і лінгвістики з початком впровадження перших обчислювальних машин, а перші експерименти в цій сфері були пов'язані з машинним перекладом на початку 50-х рр. XX століття.

Цей напрямок став активно розроблятися в 60-70 рр., і під ним у першу чергу розумілося використання статичних методів у мовознавстві, звідси і назва «Computational Linguistics» (у букв. перекл. «обчислювальна лінгвістика»). В Україні родинні терміни «обчислювальна лінгвістика», «математична лінгвістика» одержали поширення в 70-х роках XX століття. Наприкінці XX століття у зв'язку з розвитком комп'ютерних технологій та їхніх активних додатків у лінгвістичних завданнях, цей термін як назва науки трансформувався, і наука одержала більш чітке найменування «комп'ютерна лінгвістика».

Більшу організаційну та наукову працю проводить міжнародний орган – Асоціація комп'ютерної лінгвістики, що має регіональні структури в декількох країнах світу.

З погляду сучасного підходу основним напрямком комп'ютерної лінгвістики є Natural Language Processing (NLP), або завдання автоматичної обробки мови, що включає завдання аналізу та моделювання мовної структури, а саме:

- графематичний/фонематичний аналіз мови;
- морфологічний аналіз;
- лексико-граматичний аналіз мови;
- синтаксичний аналіз, або парсінг;
- аналіз і моделювання семантичної структури;
- завдання синтезу мовних елементів, у т.ч. генерація текстів.

Лінгвісти також включають наступні прикладні напрямки в область інтересів комп'ютерної лінгвістики. Більшість нижченаведених завдань пов'язано із проблемами автоматичної обробки тексту:

- машинний переклад;
- розпізнавання та синтез мовлення;
- лінгвістичні основи інформаційного пошуку;
- автоматичне індексування, реферування і класифікація текстів;
- автоматичний контент-аналіз тексту;
- авторизація текстів;

- мережні технології подання тексту та інформації на ПМ;
- корпусна лінгвістика.

Також до завдань комп'ютерної лінгвістики ми відносимо:

- розробку і використання штучних мов, у тому числі мов програмування, мов інформаційних систем;
- комп'ютерну лексикографію і термінографію;
- комп'ютерну лінгводидактику.

Окремо в рамках комп'ютерної лінгвістики звичайно виділяють завдання і технології використання статистичної лінгвістичної інформації, або автоматичну лінгвостатистику.

Таким чином, Комп'ютерна лінгвістика (Computational Linguistics), будучи одним з напрямків прикладної лінгвістики, вивчає лінгвістичні основи інформатики і всі аспекти зв'язку мови та мислення, моделювання мови та мислення в комп'ютерному середовищі за допомогою комп'ютерних програм, а її інтереси лежать в області:

- оптимізації комунікації на основі лінгвістичних знань;
- створення природно-мовного інтерфейсу і технологій розуміння мови для спілкування людини з машиною (це одна з основних проблем Штучного Інтелекту);
- створення і моделювання інформаційних комп'ютерних систем.

#### Заяпитання для самоперевірки

1. Встановити історичне занчення семіотики.
2. Виявити основні аспекти (або „плани”) знаку.
3. Охарактеризувати класичне розділення знаків за Пірсом.
4. Охарактеризувати розділення знаків за Ю.Лотманом.
5. Виявити знакові одиниці мови.
6. Визначити аспекти знаку в лінгвістиці.
7. Охарактеризувати основні поняття в інформатиці.

Література: [1-3, 6-7, 9-11, 13, 16, 19, 24, 31].

### Лекція № 3

Тема 3. Штучні мови. Формальна граматика – 2 год.

Мета лекції: визначити загальне місце та класифікацію штучних мов. Ознайомитись з формальною граматиною та метамовами.

### План лекції

1. Штучні мови як знакові системи.
2. Мови програмування як штучні мови.
3. Формальні методи опису штучних мов. Формальна граMATика. Поняття метамови.
4. БНФ-нотації. Синтаксичні діаграми.

### Основний зміст

#### 1. Штучні мови як знакові системи.

У попередній лекції ми говорили про лінгвістичні основи інформатики і про те, що у зв'язку з неможливістю повної формалізації ПМ для комунікації в комп'ютерних середовищах необхідна розробка спеціальних штучних мов (ШМ), що входить у сферу інтересів лінгвістики.

Штучні мови – це знакові системи, створювані для використання в тих галузях науки і техніки, де застосування природної мови обмежено, менш ефективно або неможливо.

Будь-яка штучна мова в порівнянні із природною завжди обмежена (за словником, синтаксисом, семантикою слів) і служить для рішення певних завдань.

Класифікація штучних мов:

1. Неспеціалізовані мови загального призначення (наприклад, есперанто, волапюк);
2. Спеціалізовані мови різного призначення (наприклад, символічні мови наук (математика, логіка, хімія, фізика)).

У 2-й клас входять мови людино-машинного (комп'ютерного) спілкування і реалізації комп'ютерних та інформаційних технологій (мови програмування, мови операційних систем, мови інформаційних систем і т.ін.).

Лінгвіст повинен мати уявлення про те, які бувають мови, яка структурна організація мов та як створюються штучні мови. Цьому і присвячені наступні наші міркування.

## 2. Мови програмування як штучні мови.

Мови програмування (МП) – це клас штучних мов, призначених для обробки інформації за допомогою комп'ютера. Будь-яка мова програмування – це строга (формальна) знакова система, за допомогою якої записуються комп'ютерні програми. За різними оцінками, на цей час існує від тисячі до десяти тисяч різних мов програмування.

Історично мови програмування виникли в 40-х роках ХХ ст. та якісно вдосконалювалися у бік спрощення свого опису, тобто високорівневої організації, методів програмування та додатка для обробки інформації людиною.

Можна виділити наступні якісні рівні розвитку МП (тобто те, як їх класифікують у програмуванні):

- низький рівень (робота з машинними кодами, наприклад, є мови – асемблери – це версії машинних кодів, адаптованих під апаратні платформи комп'ютерів);

- *середній рівень*.

- високий рівень (мова програмування високого рівня – це мова, команди і структура якої зручні для сприйняття людиною, наприклад, Паскаль, Delphi, C++, Java, SQL, JavaScript, PHP та ін.)

Якщо глянути на мову програмування як на якийсь об'єкт із лінгвістичної точки зору, можна побачити, що МП так само, як і будь-яка мова, має свої яруси або структуру.

Перший нижчий рівень – символний, його елементи алфавіту – букви, спецсимволи (за аналогією з ПМ – графематичний рівень).

Другий рівень – це рівень імен, наприклад, зарезервованих слів, виразів (в ПМ – це лексичний рівень).

Третій рівень – операторний (командний), де синтаксичні конструкції мають наказовий характер (в ПМ – аналог синтаксичного рівня), і останній – рівень програми, що завжди є синтаксично і семантично закінченою послідовністю приписань-

команд. Програма – це структурно строгий текст, записаний за формально заданими правилами штучної мови програмування.

### *3. Формальні методи опису штучних мов. Формальна граматики. Поняття метамови.*

Загальною ознакою опису спеціалізованих штучних мов є формальний метод їхнього опису й визначення шляхом завдання алфавіту, словника та системи правил утворення і перетворення виразів (граматика). Формальний метод служить для породження «правильних виразів» («правильних» – значить «записаних за певними правилами»).

Наприклад, для мов програмування задаються певні форми мовних елементів (алфавіт, слова), правила побудови команд, текстів програм, тобто точно описуються семантика і синтаксис для однозначного розуміння програм комп'ютером.

Взагалі, при написанні «правильних виразів», тобто для формального опису синтаксису елементів будь-якої мови, широко використовуються такі нотації як формальні граматики.

Формальна граматики – це система строгих (часто математичних) правил, що дозволяє за допомогою однакових процедур одержувати (виводити) правильні вирази даної мови або аналізувати наявні вирази на предмет їхньої відповідності правилам мови.

Питаннями формальних грамастик і теорією формальних мов займається такий розділ мовознавства як математична лінгвістика. Вона є суміжним напрямком прикладної лінгвістики, що тісно стикається з математикою та інформатикою.

В 1957 році у своїй книзі «Syntactic structures» американський вчений-лінгвіст Ноам Хомський запропонував класифікацію формальних мов за типом правил формальної граматики.

Існує безліч видів формальних грамастик, як наприклад:

1. Регулярна граматики.
2. Контекстно-вільна граматики.
3. Граматики безпосередньо-складових.
4. Лексико-функціональна граматики.
5. Граматики Монтегію.

Коротко опишемо граматику, що породжує.

Породжуюча формальна граMATика – це система

$$G = \langle V_T, V_{HT}, S, R \rangle,$$

де  $G$  – граMATика;

$V_T$  – множина термінальних (кінцевих) символів мови;

$V_{HT}$  – множина нетермінальних символів (з яких можна виводити далі), беруться нами в прикладі нижче в кутові дужки  $\langle \dots \rangle$ ;

$S$  – початковий символ нетермінальної множини;

$R$  – система правил висновку типу  $X \rightarrow Y$  (де  $X, Y$  – ланцюжки символів з  $V_T, V_{HT}$ ).

Множина ланцюжків, виведених через  $G$  з її початкового символу  $S$ , є вирази мови, породжувані цією граMATикою  $G$  (тобто висновок ланцюжків завжди починається з нетермінала  $S$ ).

Приклад:

Формальна система: $G = \langle \{I, We, They, like, music.\}, \{S, Pr, V, N\}, S, R \rangle$ , де $\{I, We, They, like, music.\} - V_T$ $\{S, Pr, V, N\} - V_{HT}$	Система правил $R$ : $\langle S \rangle \rightarrow \langle Pr \rangle \langle V \rangle \langle N \rangle$ . $\langle Pr \rangle \rightarrow I \mid We \mid They$ $\langle V \rangle \rightarrow like$ $\langle N \rangle \rightarrow music$	Вирази, що породжуються відповідно синтаксичним правилам $R$ : <b>I like music.</b> <b>We like music.</b> <b>They like music.</b>
---	---	--

Формальна граMATика, викладена за подібними правилами, у свою чергу, працює на базі метамови, тобто спеціальної допоміжної системи знаків (нотації) для роботи з кінцевою мовою.

#### 4. БНФ-нотації. Синтаксичні діаграми.

На практиці застосовується ще одна метамова, що навіть вважають синонімічним записом або функціональним аналогом нотації формальних граMATик. Це Бекус-Наурові форми (БНФ-форми або БНФ-нотації), які, як і формальна граMATика, служать для завдання правил одержання правильних виражень і текстів (див. рис. 3.1).

Аналогічна метамова, що має графічне наочне подання – це синтаксичні діаграми, які використовуються часто при викладанні програмування на мовах високого рівня. Синтаксична діаграма – це

схема, що пояснює правило побудови або деякого елемента, виразу, або тексту (див. рис. 3.2).

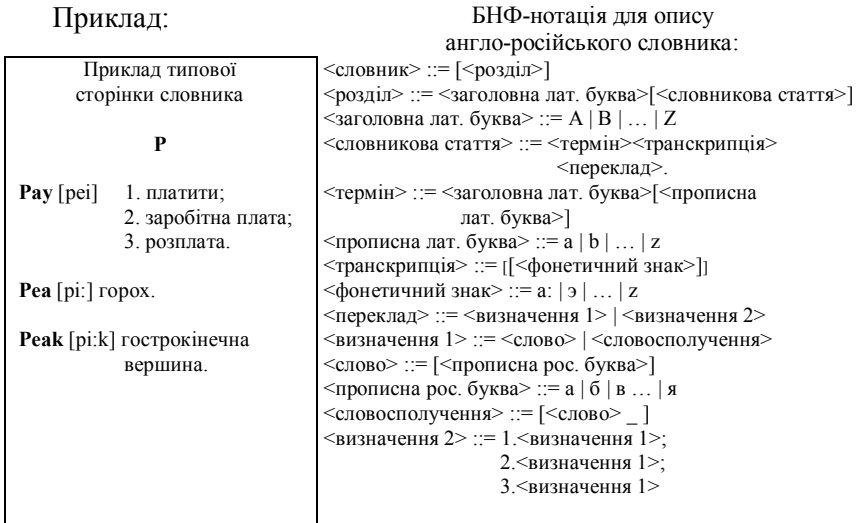


Рис. 3.1.



Рис. 3.2.

Обидві ці форми нотацій знайшли широке застосування при описі мов програмування в інформатиці. Для прикладної лінгвістики побудова формальних граматики, БНФ і синтаксичних діаграм цікава як спосіб зрозуміти структуру будь-якої мови, побачити можливості моделювання штучних мов і лінгвістичних структур.



### Запитання для самоперевірки

1. Дати загальну характеристику класифікації штучних мов.
2. Охарактеризувати види формальних граматик.
3. Визначити роль і місце метамови.

Література: [1, 4, 6-7, 9, 13, 16, 21, 23, 28, 30].

### Лекція № 4

Тема 4. Лінгвістична модель та моделі подання знань – 2 год.

Мета лекції: ознайомитись з моделюванням в лінгвістиці та моделями подання знань.

#### План лекції

1. *Моделювання та модель в лінгвістиці.*
2. *Моделі знань в штучному інтелекті. Мови подання знань як варіанти штучних мов.*

#### Основний зміст

##### 1. *Моделювання та модель в лінгвістиці.*

У різних науках доводиться мати справи з різними моделями (зразками) тих або інших процесів або об'єктів дослідження цих наук. Метод моделювання – центральний дослідницький метод практично в будь-якій науці, націлений на з'ясування способів, правильності функціонування об'єкта, або його властивостей за допомогою побудови моделі цього об'єкта. Метод моделювання мови та мовних процесів широко використовується лінгвістами, тому що дає можливість реалізувати теоретичні знання на практиці. Прикладна лінгвістика прагне будувати моделі, що відображають конкретні лінгвістичні об'єкти, їхньої системи, а також процеси мовнорозумової діяльності людини.

Поняття лінгвістичної моделі також активно використовується в комп'ютерній лінгвістиці, тому що без побудови лінгвістичних моделей неможливо вирішити багато практичних завдань створення

та використання лінгвістичних ресурсів (наприклад, машинних перекладачів, електронних словників) і завдання автоматичної обробки мови.

Усяка модель будується на основі гіпотези про можливий пристрій оригіналу і найчастіше являє собою функціональний аналог оригіналу, що дозволяє переносити отримані знання з моделі на оригінал. Критерієм адекватності моделі часто є експеримент.

Навіть у лінгвістиці, не говорячи про інші науки, існує безліч визначень поняття «модель». Найчастіше під моделлю розуміють формальний спосіб подання, опису, лінгвістичних об'єктів, наприклад:

– моделі-схеми/опису для зручного подання структури та змісту мовних об'єктів (модель словникової статті словника або див., наприклад, синтаксичну діаграму нижче).

Приклад синтаксичної діаграми – дерева залежностей (побудовано на базі Національного корпусу російської мови)



– модель-тип (language pattern) яких-небудь стійких типових одиниць досліджуваної мови (сполучуваностей слів, синтаксичних структур). Такі стійкі зразки ми неодноразово зустрічаємо в граматиках при вивченні іноземних мов. Наприклад, це стандартна для англійської мови структура речення, що називають SVO або Subject-Verb-Object pattern (ін. The Big Brother is watching you).

– модель-теорія – опис на базі строгої формалізованої наукової теорії, тобто структура елементів і правил з фіксованою метамовою (наприклад, теорія формальних граматик).

Таким чином, модель у лінгвістиці – це формалізована структура або опис із фіксованою метамовою, що служить зразком для

дослідження породжуваної мови, аналізу його характеристик і функцій. Модель завжди припускає наявність однозначно заданих об'єктів, що зв'язують їх відносини та правила звернення з ними (див. приклад формальної граматики в лекції 3).

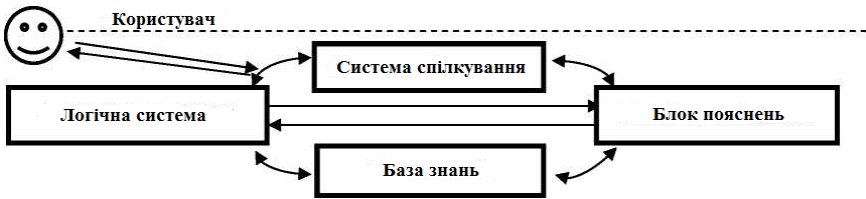
Поняття лінгвістичної моделі виникло в структурній лінгвістиці, але ввійшло в активний науковий побут в 60-70 р. ХХ ст. з розвитком математичної лінгвістики і проникненням у лінгвістику точних формальних методів дослідження. Лінгвісти звернули особливу увагу на семантику одиниць мови та на моделі мовної діяльності, тісно пов'язані з моделями мислення (які часто досліджуються в логіці). В 70-х роках ХХ ст. у результаті невдалих спроб всебічного повного моделювання природних мов у рамках рішення завдань машинного перекладу «вчені прийшли до висновку, що рішення багатьох прикладних проблем не може бути чисто лінгвістичним, а лежить на зовсім інших шляхах, на шляхах моделювання поведінки і мислення людини, семантики, синтезу формальних і семантичних засобів мови. Так з'явилася одна з найважливіших міжгалузевих фундаментальних проблем прикладної спрямованості – проблема моделювання знань».

*2. Моделі знань в штучному інтелекті. Мови подання знань як варіанти штучних мов.*

Моделювання на базі комп'ютерних технологій людських знань, розуміння мови і людського мислення – це завдання штучного інтелекту (Artificial Intelligence) як одного з провідних наукових напрямків комп'ютерної науки – інформатики, що займається створенням інтелектуальних когнітивних технологій і машин, здатних розуміти, моделювати та аналізувати тексти, зберігати і переробляти природно-мовну інформацію, приймати інтелектуальні рішення.

Система «штучного інтелекту» повинна розуміти питання людини, вирішувати інтелектуальні завдання і вести діалог з людиною на основі закладених в ній процедурних і декларативних знань. Прикладом інтелектуальної штучної системи є експертна система, якість якої визначається в першу чергу тим, наскільки природне спілкування з нею людини при рішенні завдань.

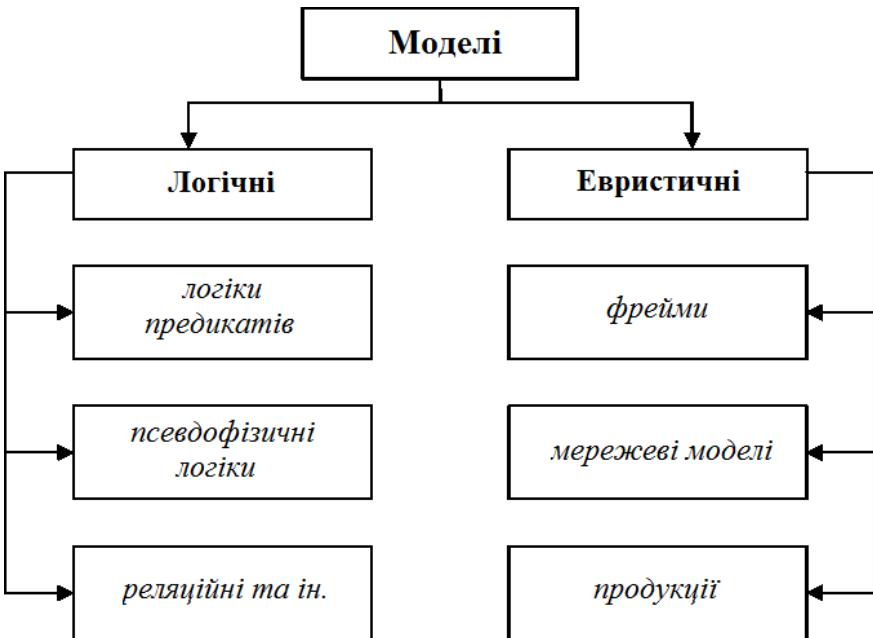
## Структура експертної системи



У штучному інтелекті важливе поняття «знання», а при побудові інтелектуальних систем – *мови подання знань*.

Знання, як правило, містять у собі 3 складові – досвід, навички та вміння – і бувають декларативними («знаю ЦО це») і процедурними («знаю ЯК це робити»).

Знання в інтелектуальній системі представляють спеціальними моделями на штучних мовах лінгвістичного забезпечення цієї системи.



Для прикладної лінгвістики цікаві мови подання знань і семантики (МПЗ). МПЗ – це штучні мови, побудовані за законами штучних мов. Найбільш популярні МПЗ – це логічні, мережеві, фреймові та продукційні.

а) Логічні МПЗ представляють знання у вигляді синтаксично правильних формул якої-небудь формальної логічної системи.

Логічних систем дуже багато. Розроблені навіть псевдофізичні логіки, тобто системи правил, що описують відносини реального миру, наприклад, логіки часу, причини та слідства, логіки просторових відношень. Часто використовують логіку предикатів та її мову логічних формул.

Приклади логічних формул:

1. Василь любить Марію.

Це речення на ПМ можна записати по-іншому

Любити (Василь, Марія)

або на мові логіки предикатів

$L(b, m)$

Предикатна формула може бути одномісна і багатомісна.

2. Василь не студент.                      Василь подарував Марії каблучку.

$\sim S(b)$

$P(b, m, k)$

Часто предикатна формула може бути вкладеною.

3. Теорема Піфагора на мові логічної формули буде мати вигляд:

РІВНІ [СУМА(КВАДРАТ (1-Й КАТЕТ),  
КВАДРАТ (2-Й КАТЕТ)), КВАДРАТ (ГІПОТЕНУЗА)]

або на символній метамові

$P1(P2(P3(k1), P3(k2)), P3(g))$

4. В логічних формулах часто використовуються спеціальні знаки – універсальні квантори, наприклад,  $\exists$  (існує...),  $\forall$  (для усіх...).

Існують дівчата красивіше Марії.

$\exists x (G(x) \ \& \ N(x, m))$

Клієнт завжди правий.

$\forall x (C(x) \rightarrow P(x))$

Спробуйте підібрати свої приклади про любов (позначимо її як L) природною мовою для таких формальних записів:

$$\begin{array}{ll} \exists x L(x, x) & \forall x (L(x, m) \rightarrow L(x, d)) \\ \sim \exists x \forall y L(y, x) & \sim \forall x \exists y (D(y) \& L(x, y)) \end{array}$$

А тепер спробуйте перетворити наступні два речення у формальний запис:

Не усі студенти дадуть відповідь на усі питання.

Ви зможете, якщо спробуєте.

Рішення завдань в логіці на логічній МПЗ – це логічний (дедуктивний) висновок за правилами цієї логічної системи.

б) Мережеві МПЗ як моделі використовують семантичні мережі, де вузли мережі – це які-небудь інформаційні одиниці – поняття, факти, процеси, імена, а дуги – відносини між ними. Відносини можуть бути будь-якими (тимчасові, причинно-наслідкові, більше-менше і т.ін.).

Мережеві МПЗ часто використовують для явного опису відносин у тій або іншій ситуації або для опису семантики структур.

Рішення завдань на мережевих моделях зводиться до пошуку фрагмента мережі, що збігається з даним зразком і до організації логічного висновку на семантичній мережі. Зв'язки елементів семантичної мережі можна представити у вигляді формул – записів на метамові семантичних мереж, тому що графічно можна представити наочно лише прості структури.

Приклад:

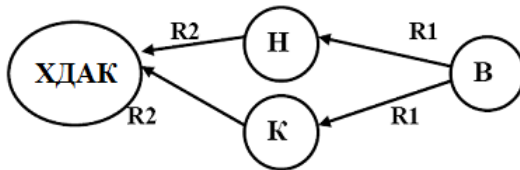
Дається наступне речення на ПМ:

«Василь бачив Наталю та Катерину у ХДАК».

Варіант семантичної мережі для цієї ситуації буде:

R1 – відношення «бачити»

R2 – відношення «знаходитись у»



Рішення завдання на мережі

Якщо у комп'ютерної інтелектуальної системи з таким знанням ситуації запитати «Василь бачив Катерину?» (тобто чи є зв'язок (B R1 K)), то вона відповість ствердливо, тому що такий зв'язок в мережі дійсно присутній. Якщо ж запитати «Катерина бачила Василя?» (K R1 B), то такий зв'язок в мережі не зазначений, і система відповість негативно.

в) Фреймові МПЗ.

Фрейм (frame – каркас, скелет, рамка) – це шаблон типової ситуації або деяка структура, що містить відомості про певний об'єкт, його характеристиках або їх значеннях і виступає як цілісна одиниця знань. Поняття «фрейму» у класичному розумінні пов'язане з мінімальним описом факту або явища, у якого не можна видалити ніяку частину опису без того, щоб не втратити повноту цього опису.

Фрейм може бути рольовий або структурний.

Структурний фрейм у більшій мері відображає декларативні знання, тобто описує структуру якого-небудь поняття, об'єкта або документа. Уявіть структуру Вашого фрейму в соціальній мережі.

Рольовий фрейм, на відміну від структурного, надає процедурні знання, наприклад, скрипти або сценарії будь-яких типових процедур або робіт. Скрипти широко використовуються в програмуванні бізнес-додатків і стандартних автоматів, наприклад, при оплаті послуг у банкоматах або для дистанційних операцій. Одна з більш складних областей використання рольових фреймів – це робототехніка.

Основні характеристики цих МПЗ – компактність і вкладеність рівнів.

У прикладі ключ – основна подія або тип сценарію, посилання – необхідна умова реалізації сценарію (послідовність дій для ключа), наслідки – результат виконання сценарію. Сценарій завершується, якщо реалізовано ключову подію (ключ) і досягнута мета діяча. R1 – відношення часу «раніше».

Приклад формального запису сценарію і модель варіанта сценарію:

БНФ-нотація сценарію:	Сценарій посадки у літак:
<p>&lt;СЦ&gt; ::= (СЦЕНАРІЙ &lt;ім'я СЦ&gt;:          &lt;ім'я слоту 1&gt; : &lt;значення слоту 1&gt;;          .....          &lt;ім'я слоту n&gt; : &lt;значення слоту n&gt;).</p> <p>&lt;ім'я СЦ&gt; ::=          &lt;слово&gt;   &lt;словосполучення&gt;          &lt;ім'я слоту&gt; ::= діяч   учасники            мета діяча   мета учасників   місце            час   посилання   наслідки   засоби          реалізації   побічні дії   ключ          &lt;значення слоту&gt; ::= «&lt;текст&gt;»            «&lt;текст&gt;» R<sub>1</sub> «&lt;текст&gt;» ... R<sub>n</sub>          «&lt;текст&gt;»</p>	<p>Сценарій посадки у літак:          (СЦЕНАРІЙ Посадка у літак:          діяч: «співробітник          авіакомпанії»;          учасники: «пасажири рейсу №          27»;          мета діяча: «zareєструвати усіх          пасажирів» R1 «посадити їх у          літак»;          мета учасників: «пройти          реєстрацію, здати багаж та сісти у          літак»;          місце: «аеропорт Київ-          Бориспіль»;          час: «впродовж 2 годин до          відльоту рейсу № 27»;          посилання: «реєстрація» R1          «збір на посадку» R1 «контроль на          митниці» R1 «подача автобуса» R1          «подача трапа»;          наслідки: «заповнення літака          пасажирами»;          ключ: «посадка у літак усіх          пасажирів, що пройшли          реєстрацію»).</p>

### г) Продукційні МПЗ

Продукції – це одна з популярних форм подання процедурних знань в експертних системах та інших системах знань, що працюють зі сценаріями.

Продукції представляють в основному процедурні знання, причому формула правил продукції відносно проста – «ЯКЩО..., ТО ...»:

if <condition> then <conclusion> або  
 if <condition> then <action>

Наприклад,

ЯКЩО <температура тіла більше 40°>,

ТО <терміново викликай швидку допомогу за телефоном 103>.



Продукції можуть бути складовими, наприклад,  
if a and b and c then d.

У формі правил-продукцій можна задавати і лінгвістичні знання про мову, отримані в результаті аналізу безлічі текстів.

Завдання для продукційної моделі ставляться як завдання пошуку потрібної послідовності продукцій, при якому досягається потрібна мета.

Запитання для самоперевірки

1. Охарактеризувати види моделей (поняття).
2. Навести приклади інтелектуальних систем.
3. Дати характеристику видам знань.
4. Охарактеризувати моделі подання знань.

Література: [1-3, 8-9, 12, 16, 21, 33-36].

## **РОЗДІЛ 2. ПРАКТИЧНИЙ АСПЕКТ ПРИКЛАДНОЇ (КОМП'ЮТЕРНОЇ) ЛІНГВІСТИКИ**

### **Лекція № 5**

Тема 5. Інформаційний пошук як засіб прикладної лінгвістики – 2 год.

Мета лекції: ознайомитись з поняттями інформаційно-пошукової системи та мови.

План лекції

1. *Інформаційний пошук.*
2. *Лінгвістика у завданнях інформаційного пошуку. Інформаційно-пошукові мови як штучні мови.*

Основний зміст

1. *Інформаційний пошук.*

Теорія і практика інформаційного пошуку стала розвиватися в середині ХХ століття. До кінця століття із впровадженням світових інформаційних комп'ютерних мереж та їхнього інформаційно-лінгвістичного забезпечення ця область стала окремим визнаним науково-практичним напрямком комп'ютерної науки зі своєю теорією інформаційного пошуку та численних практичних розробок, продуктами і технологіями.

Пошук інформації (Information Retrieval) – це процес відшукування в деякій системі зберігання інформації таких документів (текстів, записів і т.ін.), які відповідають запиту, що надійшов. У якості таких засобів зберігання і пошуку інформації виступають інформаційно-пошукові системи (IRS/Information Retrieval Systems), елементами яких є структурований масив документів (база даних/індекс), що виступають як об'єкт пошуку, різноманітні технічні та програмні засоби, як наприклад, програми-роботи, а також інформаційно-пошукова мова, що задає правила індексування документів, правила пошуку.

По В. П. Захарову «Інформаційно-пошукова система (ІПС) – це упорядкована сукупність документів (масивів документів) і інформаційних технологій, призначених для зберігання і пошуку інформації – текстів (документів) або даних (фактів). Інформаційно-пошуковими системами є будь-які певним чином організовані сховища інформації. Причому інформаційно-пошукові системи можуть бути і неавтоматизованими. Головне – це цільова функція: зберігання і пошук інформації».

При введенні документа в базу даних ІПС його індексують (тому саму базу пошукової системи часто називають індексом). Процес індексування в основному пов'язаний з визначенням і вибіркою ключових слів оброблюваних документів і вираженням їх формально у вигляді пошукового образу. Так база даних ІПС складається з безлічі індексних пошукових образів. Безпосередньо при пошуку виконується зіставлення вашого запиту, тобто того, що ви вказали в запиті з пошуковим образом, тобто тим, що зберігається в індексі.

Для більш наочного розуміння способу зберігання інформації в базі ІПС приведемо наступний простий приклад.

Приклад:

		Словник інформаційно-пошукової мови						
АДР ЕСИ доку мент ів		S1	S2	S3	S4	S5	S6	Тут: S1-S6 – дескриптори (ключове слово); ♥ – умовний знак того, що цей дескриптор знаходиться у документі; 1-5 – адреси або номери документів.
	1	♥	♥			♥		
	2			♥	♥		♥	
	3	♥	♥	♥				
	4			♥	♥			
	5					♥	♥	

Наприклад, документ №3 має ПО = {S1, S2, S3}, а документ №5 – {S5, S6}.

Якщо пошуковий припис-запит користувача буде ПП=S1 AND S2, то пошукова система видасть адреси документів №1 і №3.

Якщо пошуковий припис-запит користувача буде ПП=S5 OR S6, то пошукова система видасть адреси документів №1, №2, №5.

Залежно від об'єкта зберігання і типу запиту часто в навчальній літературі розрізняють два види інформаційного пошуку:

- документальний (коли користувач шукає який-небудь текст/документ);
- фактографічний (коли користувач шукає які-небудь фактичні дані, наприклад, «День народження Анджеліни Джолі»).

Характеристики інформаційного пошуку – це такі його семантичні показники як:

- повнота видачі інформації;
- точність її видачі;
- втрати інформації;
- інформаційний шум.

Повнотою пошуку (Recall) називається міра, що обчислюється як відношення кількості виданих релевантних документів до загального числа релевантних документів, що містяться в базі інформаційно-пошукової системи.

Точність пошуку (Precision) – це відношення кількості виданих системою релевантних документів до загального числа документів у видачі.

Розглянемо наступну таблицю, за якою легко прораховуються основні показники інформаційного пошуку:

Документи	Релевантні (необхідні, потрібні)	Нерелевантні (непотрібні)
видані ІПС	a	b
невидані ІПС	c	d

$PB = (a/(a+c)) \cdot 100\%$  – повнота видачі інформації

$VI = (c/(a+c)) \cdot 100\%$  – втрата інформації

$III = (b/(a+b)) \cdot 100\%$  – інформаційний шум

$TV = (a/(a+b)) \cdot 100\%$  – точність видачі

Особливо цікаво нам поняття «релевантність» як фундаментальне поняття теорії інформаційного пошуку. Відповідно до визначення, документ, центральний предмет або тема якого в цілому відповідає значенневому змісту інформаційного запиту, називається релевантним, а властивість значенневої близькості між документом та інформаційним запитом – релевантністю. За багатьма, найчастіше суб'єктивними, причинами релевантність була і залишається основною проблемою інформаційного пошуку.

## *2. Лінгвістика у завданнях інформаційного пошуку. Інформаційно-пошукові мови як штучні мови.*

Теорія і практика інформаційного пошуку тісно пов'язана з лінгвістикою, тому що, по-перше, основний обсяг інформації і тексти документів представлені на природних мовах, а для витягу інформації та індексування текстів необхідні знання автоматичної обробки мов, наприклад, правила комп'ютерної морфології. По-друге, інформаційні комп'ютерні системи побудовані і працюють на базі штучних мов.

З визначення інформаційно-пошукової системи видно, що її основним лінгвістичним засобом є спеціалізована штучна мова.

Інформаційно-пошукова мова (ІПМ) – це спеціалізована штучна мова, призначена для:

- опису інформаційних запитів до інформаційно-пошукових систем (мова запитів);

- опису формальних характеристик документів у вигляді пошукового образу, що зберігається в базі системи (мова індексування).

Необхідність впровадження штучних мов викликана необхідністю усунення надмірності природної мови для інформаційного пошуку, а також ліквідації мовної синонімії, омонімії і неоднозначностей різного роду.

Інформаційно-пошукова мова, як і будь-яка мова, складається з фіксованих одиниць, наприклад, має свій словник і синтаксис, і є штучною мовою, тобто обмежена за своєю формою і структурою завданнями на пошук.

Розглянемо кожний з варіантів ІПМ більш докладно:

### 2.1. Мови запитів.

Здавані простими на перший погляд мови запитів являють собою досить комплексні системи правил і процедур. Як правило, модель мови запитів включає наступні елементи:

- пошукові одиниці (ключові слова, що виражають інформаційну потребу користувача);

- засоби лематизації лінгвістичних одиниць запиту (приведення слів до нормальної словникової форми);

- спеціальні пошукові (булеві) оператори, типу OR, NOT, AND;

- засоби лінійної граматики (оператори відстані, позиційні оператори, наприклад, title:);

- додаткові умови пошуку: обмеження області пошуку по мові, регіону, даті створення документа і т.ін.

Запит на пошук найчастіше обмежений по мові і може мати певне формальне подання. Самий частотний спосіб пошукового запиту – через обмежений набір ключових слів, що задається користувачем у пошуковому рядку. Наприклад, знайти будь-яку книгу в Інтернеті можна, задавши ключові слова «Автор назва книги» в пошукових системах Яндекс, Google і т.ін.

Окрім стандартної практики в пошукових системах передбачена функція розширеної мови запитів, її можна легко знайти на сайтах пошукових машин – <http://yandex.ru/search/advanced> або [http://www.google.ru/advanced\\_search](http://www.google.ru/advanced_search). Розширена пошукова мова

включає множину операторів для поліпшення якості та звуження зони пошуку релевантної інформації.

Наприклад:

- ”чемпіонат світу 2018” OR ”олімпійські ігри 2018”;
- мумій троль мультфільм -рок -лагутенко – Пошук у Яндексі або Google тільки мультфільму, а НЕ (-заперечення) групи Іллі Лагутенко;
- соціолінгвістика -site:ru.wikipedia.org (шукаємо інформацію скрізь крім сайту російської Вікіпедії);
- date:ГГГГ{\*|ММ{\*|ДД}} – Пошук у Яндексі тільки по сторінках, дата яких задовольняє заданій умові (наприклад, шукаємо праці професора ХДАК тільки за 2012 р. – Шейко В. М. date: 2012).

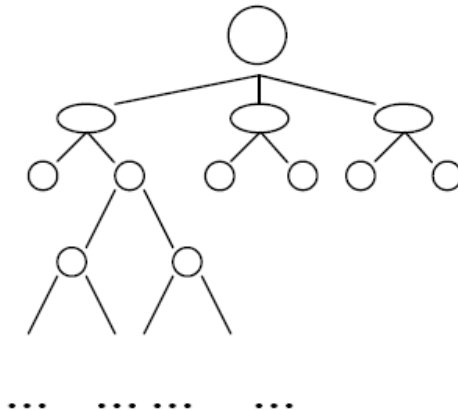
## 2.2. Мови індексування

У літературі по теорії інформаційного пошуку звичайно виділяють наступні розповсюджені види інформаційно-пошукових мов для індексування документів:

1. Мови класифікацій – ієрархічні; алфавітно-предметний та ін.;
2. Дескрипторні мови.

Ієрархічні ПІМ (ієрархія – класифікація від загального до окремого). Така організація використовується для пошуку книг у бібліотеці (наприклад, Універсальна десяткова класифікація – УДК).

Дерево класів



Алфавітно-предметні ІІМ представляються як алфавітний список ключових слів якого-небудь документа з позначками (наприклад, алфавітно-предметний покажчик наприкінці книг). Використовується для побудови різних покажчиків, каталогів, картотек.

Дескрипторна мова – найбільш природна і популярна форма мов індексування для вираження пошукових образів і у теперішній час широко використовується в сучасних інформаційно-пошукових системах (зокрема в багатьох пошукових системах мережі Internet).

У таких мовах використовується принцип координатного індексування, тобто перераховуються ключові слова або дескриптори, які виражають центральну тему або цілісну характеристику шуканого об'єкта. При цьому використовується принцип логічного множення понять, у результаті якого із простих лексичних одиниць будуються більш складні, що виражають більш вузькі поняття. Цей принцип опису змісту документів через перерахування ключових слів існує здавна. Наприклад, перетинанням понять ПАРИЖ і ПАМ'ЯТКИ, заданих у запиті, породжується нове більш вузьке поняття ПАМ'ЯТКИ ПАРИЖУ.

Як лексичні одиниці в дескрипторних ІІМ виступають дескриптори – імена понять або класів понять, які явно перераховуються в дескрипторному словнику. Це слова (або словосполучення), обрані у якості представників класів і груп синонімічних слів і словосполучень. Як правило, це іменники або номінативні вираження.

Списки дескрипторів організовані в спеціальні семантичні словники (пошукові тезауруси). Пошукові тезауруси (одномовні або багатомовні) – спеціальні словники для інформаційного пошуку по масивах природно-мовних документів, організовані за принципом зіставлення слів з їхніми поняттями. Їхня структура і розробка часто стандартизуються.

Такий дескрипторний словник використовується як засіб лексичного контролю (наприклад, зняття омонімії, синонімії одиниць) при індексуванні документів і запитів. Наприклад, при індексуванні всі синоніми запиту і пошукового образу представляються однією і тією ж лексичною одиницею – дескриптором (порівн. лінгвістика – мовознавство, наука про мову).

Структурно пошуковий тезаурус являє собою алфавітний список дескрипторів разом з їхніми словниковими статтями (гніздами). Словникова стаття звичайно містить:

- заголовний дескриптор;
- ключові слова або словосполучення, що входять у гніздо даного дескриптора (умовні синоніми);
- «вищестоящі» дескриптори (перебувають з даним у відношенні «рід-вид», «частина-ціле»);
- «нижчестоящі» дескриптори (перебувають із заголовним дескриптором у відношенні «вид-рід», «ціле-частина»);
- асоціативні дескриптори (пов'язані з даним іншими різноманітними відносинами: причина-наслідок, процес-об'єкт, властивість-носій властивості, функціональна схожість).

Зазначені відносини приводяться з позначками, найчастіше так:

Словникова стаття інформаційно-пошукового тезаурусу

**ДЕСКРИПТОР**

с – ключові слова (синоніми)

в – родові слова

(дескриптор, що підпорядковує даний)

н – видові слова

(дескриптор, підпорядкований даному)

а – асоціації (відношення)

Потрібно відзначити, що тезауруси бувають не тільки інформаційно-пошукові, а їх структура не обмежується тільки представленими відносинами. Тезауруси відносяться до дуже цікавого класу семантичних словників, що ми розглянемо в наступній лекції, присвяченій одному з древніх практичних напрямків лінгвістики – лексикографії.

Запитання для самоперевірки

1. Дати загальну характеристику поняття інформаційно-пошукової системи та видів інформаційного пошуку.
  2. Охарактеризувати процес індексування бази даних.
  3. Визначити роль і місце інформаційно-пошукової мови.
- Література: [1-3, 5, 9, 11, 24, 28, 34].



## Лекція № 6

Тема 6. Лексикографія та словники – 2 год.

Мета лекції: визначити роль лексикографії та словника.

### План лекції

*1. Лексикографія як один з важливих напрямків прикладної лінгвістики. Традиційна та машинна (комп'ютерна) лексикографія. Основні напрямки комп'ютерної лексикографії.*

*2. Словник як об'єкт лексикографії. Класифікація та організація словників.*

*3. Ідеографічні словники і тезауруси як приклади словників.*

### Основний зміст

*1. Лексикографія як один з важливих напрямків прикладної лінгвістики. Традиційна та машинна (комп'ютерна) лексикографія. Основні напрямки комп'ютерної лексикографії.*

Лексикографія (від грецьк. *lexis* ‘слово’ та *grafia* ‘писання, наука’), будучи одним з важливих напрямків прикладної лінгвістики, займається теорією і практикою складання словників.

У лексикографії виділяють два науково-практичних напрямки: традиційна і машинна (комп'ютерна) лексикографія.

Традиційна лексикографія має глибокі історичні коріння і в більшій мірі займається теорією та практикою складання «традиційних» словників. Самі ранні глоси (від грецького *glossa* ‘мова, слово’, словникові позначки про значення незнайомих слів) відомі з найглибшої стародавності (наприклад, шумерські глоси XXV ст. до н.е.). Тому, найпершими типами словників багато вчених вважають *гласарії*, написані від руки списки іноземних і незвичайних слів, з якими доводилося стикатися в манускриптах на древніх мовах.

Машинна лексикографія (*Computational Lexicography*) займається автоматизацією підготовки словників і вирішує завдання

розробки електронних словників. На відміну від традиційної, машинна (комп'ютерна) лексикографія – відносно молода наука, що реалізує традиційні наробітки в технічних середовищах та створює різноманітні електронні словники. До основних завдань комп'ютерної лексикографії відносяться також завдання розробки технологій складання електронних словників та управління термінологією (Terminology Management).

У цей час можна виділити три основних напрямки комп'ютерної лексикографії:

1) автоматичне одержання з тексту різних словників (наприклад, термінологічних, частотних словників, словників конкордансів та ін.);

2) створення словників, що є електронними версіями традиційних словників (наприклад, словник Даля), або комплексних електронних лінгвістичних словників для традиційних словникових робіт, наприклад, відомий словник LINGVO [<http://lingvopro.abbyyonline.com/ru>]. Великий вибір такого роду відкритих словників доступний через пошукові системи, наприклад, <http://slovari.yandex.ru/>.

3) розробка теоретичних і практичних аспектів складання спеціальних комп'ютерних словників, наприклад, для інформаційного пошуку, машинного перекладу (наприклад, пошукові тезауруси, словники стоп-слів, словники основ та флексій для морфологічного аналізу/синтезу).

Таким чином, ми бачимо, що основним об'єктом інтересів лексикографії є такий продукт як словник і всі завдання, пов'язані з розробкою різного типу словників.

## *2. Словник як об'єкт лексикографії. Класифікація та організація словників.*

Словник – певним чином організована збірка слів, звичайно із приписаними їм коментарями, в яких описуються особливості їхньої структури та/або функціонування. Крім слів, об'єктами словникового опису можуть виступати їхні компоненти (такі, наприклад, словники морфем), словосполучення різних типів, стійкі групи – прислів'я, приказки, цитати і т.ін. В іншому значенні термін

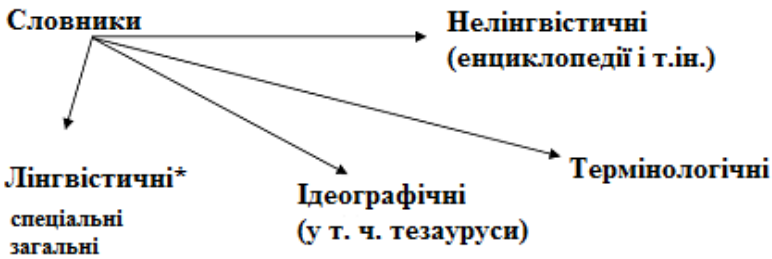
«словник», або лексикон, позначає всю сукупність слів деякої мови (інакше кажучи, його лексику) і протиставляється терміну «граматика», що позначає сукупність правил побудови зі слів більш складних мовних виражень.

Таким чином, під словником розуміють:

- повний словниковий склад мови;
- упорядкована для рішення практичних завдань множина лексичних одиниць;
- довідкову книгу слів, розташованих у певному порядку, що дає кому-небудь інформацію про те чи інше слово.

Основне завдання словника – це подання або опис лексики мови та її особливостей для рішення конкретних завдань. Це найскладніша проблема, тому що лексика мови має тенденцію збільшуватися і якісно змінюватися.

Класифікацію словників схематично можна представити в такий спосіб.



\* Примітка: загальні лінгвістичні словники (тлумачні, перекладні, орфографічні, частотні та ін.), спеціальні (синонімів, антонімів, словник прислів'їв і приказок, сленгу, словники іноземних слів та ін.).

Основною структурною одиницею словника (як книги) є словникова стаття, організація і моделювання якої є найчастіше складною прикладною проблемою та завданням лексикографії.

Реєстр слів словника – це перелік термінів словника без їхніх тлумачень.

Важливим питанням при складанні словників також є порядок розташування словникових статей, найчастіше це алфавітний порядок або предметний (тематичний), при якому слова групуються

за темами або графічно (наприклад, тематичний візуальний словник).

Електронні словники часто представляють із себе складний комплекс комп'ютерних програм – лінгвістичних платформ. Прикладом тому служить АBBYY Lingvo Content – професійний додаток, що дозволяє легко створювати нові словники і глосарії, редагувати їх і публікувати в паперовому та електронному вигляді на сайтах, на корпоративних порталах, а також у вигляді додатків для PC, Mac, смартфонів і мобільних пристроїв ([http://www.abbyy.ru/lingvo\\_content](http://www.abbyy.ru/lingvo_content)).

Відзначимо деякі особливості автоматичних словників:

- крім словникової бази даних (перелік слів за алфавітом) для автоматичної словникової роботи необхідні спеціальні алгоритми і програми, наприклад, морфологічної нормалізації, або лематизації. Лематизація – це приведення різних форм слова до його канонічної (вихідної) форми (одне із завдань комп'ютерної морфології);

- у машинних словниках присутні не тільки переліки окремих слів, але і до 50% словосполучень (особливо в термінологічних словниках, які дуже важливі для перекладу спеціальних і технічних текстів);

- електронні словники одержали в цей час велике поширення в силу їхньої доступності для широкого кола користувачів через різні мобільні і персональні пристрої.

Примітка: варто відізнати автоматичні словники і системи машинного перекладу (як не дивно, їх часто плутають обивателі). Останні лише включають перші, будучи значно складніше.

### *3. Ідеографічні словники і тезауруси як приклади словників.*

Ми вже згадували словник-тезаурус як тип семантичного словника в попередній лекції. Однак тезауруси мають більш широкі додатки, ніж ми вказали в лекції, присвяченій інформаційному пошуку.

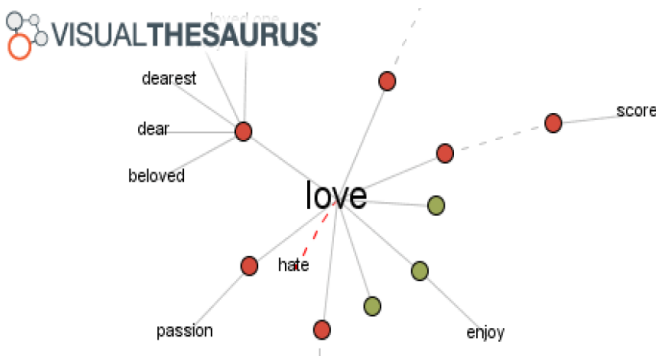
Клас ідеографічних словників (предметні, тематичні), до яких часто відносять і тезауруси, – це особливого роду словники, організовані, по-перше, за тематичним принципом, і, по-друге, за принципом ієрархії відносин і «від змісту до слова», тобто ідеографічні словники орієнтовані на семантику мови, і кожен

такий словник – це деяка семантична модель лексики, побудована на ієрархічних відносинах типу «рід-вид», «частина-ціле», «синоніми» і т.ін.

Тезаурус – це також ідеографічний словник, але має чітку складну ієрархію відносин. У словникових статтях тезаурусу відбиті істотні для даного терміну зв'язки з іншими поняттями, іншими словами – це маленька енциклопедія.

Найвідоміший класичний словник-тезаурус – це тезаурус П. Роже «Thesaurus of English words and phrases», опублікований в 1852 р., що неодноразово перевидавалася як класична лексикографічна праця і представлений у мережі Інтернет. П. Роже систематизував лексику англійської мови за категоріальними групами, кожна з яких представлена ім'ям поняття, яких спочатку налічувалося 1000. Слова розташовані за абеткою, далі йдуть синоніми слів з частинами мовлення (іменники, дієслова, прикметники, прислівники), антоніми і потім списки ріднинних слів. Багатьма відзначається, що цінність цього тезаурусу в його природності, у тім, що цей опис загальної понятійної лексики мови, а також у тім, що його можна залучати до використання в комп'ютерних системах як семантичний засіб.

Напрямок, пов'язаний з розробкою семантичних словників, активно розвивається. Існують глобальні проекти семантичних словникових баз типу WORDNET, а також такі цікаві версії словників як візуальні тезауруси, що наочно представляють візуальні карти обраних слів, наприклад:



## Запитання для самоперевірки

1. Дати загальну характеристику напрямків лексикографії.
  2. Охарактеризувати напрямки комп'ютерної лексикографії.
  3. Дати загальну характеристику лінгвістичним платформам електронних словників.
  4. Визначити роль і місце тезаурусу.
- Література: [3-4, 9, 16, 11, 21, 29-35].

**Лекція № 7**

Тема 7. Машинний переклад та технології – 2 год.

Мета лекції: визначити загальне місце систем машинного перекладу та технологій машинного перекладу.

## План лекції

1. *Машинний переклад. Види МП.*
2. *Технології МП. Завдання машинного перекладу як одне з важливіших завдань прикладної лінгвістики. Основні етапи системи МП, що працює на базі правил.*

## Основний зміст

*1. Машинний переклад. Види МП.*

Переклад з однієї мови на іншій у загальному випадку складається в зміні алфавіту, лексики і синтаксису мови зі збереженням його семантики.

Переклад – це вид інформаційної діяльності, потреба в якій ніколи не скорочується, а навпаки збільшується. Дослідження ринку перекладів показали, що обсяги цього виду діяльності постійно збільшуються, а в складі перекладів переважають спеціальні, науково-технічні переклади – до половини загального обсягу ринку перекладів, потім йдуть усний, навчальний, синхронний, ..., художній.

Проблема моделювання перекладу для додатка її в комп'ютерному середовищі є центральною проблемою як прикладної лінгвістики, так і штучного інтелекту. Очевидно, що автоматизація перекладу дозволить підвищити його ефективність, а також розширить границі міжлюдської комунікації.

Машинний переклад – це перетворення комп'ютером тексту на одній природній мові в еквівалентний за змістом текст на іншій природній мові.

Системи МП – це цілий комплекс спеціальних найскладніших програм та алгоритмів плюс спеціальні автоматичні словники і формалізовані граматики для кожної мовної пари (вхідної і вихідної мов).

Історія цього прикладного напрямку досить добре задокументована в багатьох джерелах. Історію МП відлічують з 1946 р., з виникнення такої науки, як кібернетика (наука про управління). В 1954 р. був проведений так званий Джорджтаунський експеримент: переводився текст (250 слів) з російської на англійську мову. Перша промислова система МП SYSTRAN починає функціонувати в США в 1970 р. Перші досвіди розробки комерційних систем по МП показали величезні труднощі при моделюванні семантики мовних одиниць і дозволі лінгвістичних неоднозначностей різного роду. Проблеми, пов'язані з моделюванням всієї структури мови, повністю не вирішені дотепер, тому були запропоновані інші підходи до організації машинного перекладу, що працюють на базі інформації з більших обсягів текстів, наприклад, так з'явилися технології CAT – Computer-Aided-Translation, а також статистичний переклад на базі пошукових систем, наприклад, сервіс Google Translate, що працює по власних алгоритмах з 2007 р.

Існує кілька підходів до класифікації систем машинного перекладу, виходячи з різних критеріїв.

Наприклад, їх часто ділять залежно від ступеня участі людини в процесі перекладу. Відповідно до цього критерію всі системи діляться на три типи:

- FAMT (Fully-Automated Machine Translation) – повністю автоматичний машинний переклад;

- HAMT (Human-Aided Machine Translation) – машинний переклад за участю людини;

– МАНТ (Machine-Aided Human Translation) – переклад, здійснюваний людиною із залученням допоміжних програмних і лінгвістичних засобів.

Види машинного перекладу:

1) інформативний МП – грубий послівний переклад без участі людини, достатній для поверхневого ознайомлення зі змістом тексту (автоматичний). Активно експлуатується і зараз при пошуку інформації в пошукових системах;

2) професійний МП – більш якісний переклад на базі комерційних систем МП із наступним редагуванням людиною (автоматизований);

3) інтерактивний (персональний) МП – вважається перекладом у спеціальних системах підтримки, наприклад, у таких як CAT, що експлуатує бази перекладу (Translation Memory), проходить у режимі діалогу людини з комп'ютерною системою. Активно використовується комерційними компаніями.

Завдяки постійному розвитку технологій, нагромадженню лінгвістичних ресурсів і лінгвістичної статистики більшість сучасних систем МП дає непогана якість перекладу. Якість МП багато в чому залежить від можливостей настроювання, наявних ресурсів (наприклад, наявності готових перекладацьких баз), від типу текстів (очевидно, що художні тексти не має змісту перекладати за допомогою систем МП, а тексти інструкцій до пристроїв переводяться з гарною якістю навіть в автоматичному режимі).

*2. Технології МП. Завдання машинного перекладу як одне з важливіших завдань прикладної лінгвістики. Основні етапи системи МП, що працює на базі правил.*

Як правило, виділяють наступні технології, на базі яких працюють сучасні системи машинного перекладу:

1. Example-based Machine Translation, або Машинний переклад на базі готових прикладів перекладів.

Ця концепція МП була запропонована японським дослідником М. Нагао в 80-х роках ХХ століття. Ідея полягала в наступному – при нагромадженні досить великої колекції раніше переведених



фраз для конкретних типів текстів та у вузьких областях велика ймовірність того, що більша частина наступних текстів буде аналогічна вже переведеним вручну. Яскравим представником цього підходу є інструменти CAT (Computer-Aided-Translation tools), що реалізують технології автоматизованої підтримки перекладу.

До систем CAT відносять ті системи, які забезпечують роботу на базі «пам'яті перекладів» (Translation Memory), тому їх часто називають ТМ-інструменти. Функція звертання до баз «пам'яті перекладу» дає перекладачеві і компанії безсумнівні переваги:

- однакову лінгвістичну одиницю або речення, що зустрічаються в різних місцях, не потрібно перекладати двічі. Якщо подібна відповідність переведена, то необхідно лише відкоригувати попередній переклад;

- якщо переклад тексту виконує група перекладачів, то закінчений текст виходить більш термінологічно і стилістично однорідним.

База даних ТМ складається з пар паралельно зіставлених сегментів (найчастіше це речення), бази формуються програмно по ходу перекладу (споконвічно САТ-система порожня) і зберігаються окремо. САТ-програма при перекладі робить звернення до бази, і якщо новий сегмент має точну або часткову відповідність у наявній накопиченій базі перекладів, то пропонується варіант перекладу, якщо відповідності ні, то переклад іде вручну і варіант записується в базу як новий сегмент.

Багатьма користувачами відзначається, що при використанні САТ-систем значно підвищується ефективність роботи перекладача, що працює з типовими спеціальними текстами, для яких вже накопичена багата практика перекладацьких прикладів, тобто бази перекладів.

Самими популярними є пакети SDL Trados, Deja Vu. Крім них існує велика кількість як платних, так і безкоштовних версій програм автоматизованого перекладу, функціонально програми схожі. Виходять демонстраційні безкоштовні версії програм. Ринок САТ-програм постійно розвивається, паралельно накопичується безліч спеціальних баз перекладів.

## 2. Statistical Machine Translation, або Статистичний МП.

Ця технологія машинного перекладу також експлуатує ідею вищеописаного ЕВМТ-підходу, але має більш строгу математичну базу, з огляду на статистику тих лінгвістичних закономірностей, які отримані на базі структурного аналізу текстів та аналізу доступних паралельних корпусів текстів. Корпус паралельних текстів – це тексти, що містять речення на одній мові і відповідні їм речення на другому (наприклад, двомовні сайти). Статистичний МП використовує властивість «самонавчання» мові (machine learning), тобто чим більше накопичено паралельних текстів і чим точніше вони відповідають один одному, тим краще результат статистичного машинного перекладу.

Вважається, що інтерес до використання лінгвостатистичних методів у МП виник ще наприкінці 40-х років ХХ століття паралельно з розвитком теорії інформації К. Шеннона. В 60-х роках ХХ століття дослідники успішно застосовували лінгвостатистику до рішення завдань дешифрування древніх текстів. Але активно цей напрямок став розвиватися тільки на початку 90-х років ХХ століття під час технологічного та інформаційного буму, нагромадженням даних у корпусній лінгвістиці, машинному навчанні та інформаційному пошуку. Тому глобальні пошукові системи стали впроваджувати сервіс статистичного перекладу в свої інтерфейси. Наприклад, інформаційно-пошукова система Google перейшла на цю технологію в 2007 році і пропонує сервіс Google Translate. На початку 2011 року Яндекс, як раніше Google, впровадив власну подібну систему машинного перекладу.

3. Rule-based Machine Translation, або Машинний переклад на базі лінгвістичних правил, – це класична технологія МП, що різко відрізняється від попередніх, з розробки якої і почалася історія МП та комп'ютерної лінгвістики.

Глобальним завданням такої технології машинного перекладу є формальне моделювання природних мов як об'єктів, реалізованих у певному програмному та апаратному комп'ютерному середовищі, а також моделювання різних процедур аналізу, зіставлення (трансферу) і синтезу лінгвістичних одиниць і текстів для цих природних мов. Тобто таке завдання вимагає реалізації процедур:

- 1) графематичного (символьного) аналізу;
- 2) морфологічного і лексичного аналізу та синтезу;
- 3) синтаксичного аналізу та синтезу;

## 4) семантичних перетворень.

При МП обробляються всі рівні мовної структури, переходячи від одного підзавдання до іншого. Для моделювання МП необхідне моделювання його підзавдань, розробка алгоритмів і комп'ютерних програм для роботи цих алгоритмів. Якщо уважно придивитися до цих завдань МП, то можна побачити, що вони представляють окремі та складні проблеми і напрямки прикладної лінгвістики.

За історію машинного перекладу як наукового напрямку виділилися три підходи до моделювання систем такого роду:

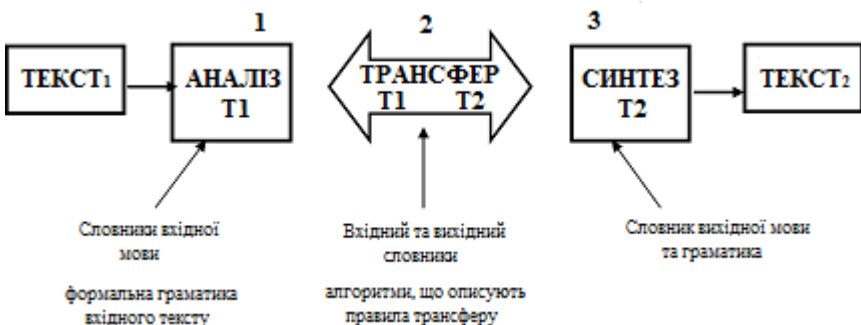
а) «прямі» системи МП давали послівний переклад (слово в слово);

б) системи «інтерлінгви», або «текст-сенс-текст», метою було розробити значеннєву мова-посередника;

Перший підхід обмежений у своїх можливостях, а для моделей-інтерлінгви поки недостатньо наукової бази та успішних розробок.

в) «текст-трансфер-текст». Переклад відбувається на рівні перекладних відповідностей мовної пари. Одиницею перекладу виступає перекладна відповідність, тісно пов'язана з лексичною одиницею (словосполучення, слово). У цій моделі є проміжний етап – трансфер, на якому відбувається встановлення перекладних відповідностей (узгодження мовних одиниць).

Класично МП розбивається на етапи аналізу вхідного тексту, трансферу та синтезу вихідного тексту.



Перетворення T1 повинне починатися з попередньої підготовки – його аналізу. Звичайно розрізняють наступні види лінгвістичного аналізу: морфологічний, лексичний, синтаксичний, семантичний.

Метою етапу аналізу є одержання всієї лінгвістичної інформації про виділені одиниці даного тексту, наприклад, одержання всієї морфологічної інформації про словоформу для наступного коректного перекладу, і побудова внутрішнього подання вхідного речення.

На етапі трансферу вже починається перекладацька робота, відбувається зіставлення одиниць по словниках, виконуються інші перетворення за допомогою спеціальних формальних правил трансферу.

Ціль етапу синтезу – на основі отриманої в результаті аналізу інформації побудувати (синтезувати) правильне речення вихідною мовою.

Ця модель побудови систем перекладу на базі правил найбільш вдала в комерційному плані і у цей час переважає над прямими та інтерлінгвальними типами МП (на ній працюють відомі промислові системи PROMT і SYSTRAN).

Були коротко описані лише базові поняття і технології сучасного машинного перекладу, але цей напрямок постійно розвивається, з'являються нові способи автоматизації та підвищення ефективності перекладацької діяльності, різні гібридні технології. Наприклад, з'являються технології типу Traduwiki (<http://traduwiki.org>) для спільної перекладацької діяльності, т.зв. «хмарні технології» (cloud-based translation), а люди продовжують дотримуватися улюбленого ними, вже відпрацьованого століттями принципу «principle of least effort».

#### Запитання для самоперевірки

1. Дати загальну характеристику понять перекладу та машинного перекладу.
2. Охарактеризувати типи систем машинного перекладу.
3. Охарактеризувати технології машинного перекладу.
4. Визначити підходи до моделювання систем машинного перекладу.

Література: [3, 6, 9, 14, 16, 22, 30, 33-36].

### Лекція № 8

Тема 8. Автоматична обробка тексту. Лінгводідактика – 2 год.

Мета лекції: визначити типи автоматичної обробки тексту.

#### План лекції

1. *Текст та гіпертекст. Завдання автоматичної обробки тексту.*
2. *Сучасна лінгводидактика.*

#### Основний зміст

1. *Текст та гіпертекст. Завдання автоматичної обробки тексту.*

Коротко розглянемо інші прикладні завдання лінгвістики, які з'явилися або одержали новий розвиток з виникненням комп'ютерної техніки.

З появою комп'ютерних мереж і нових інформаційних технологій трохи трансформувалися традиційні лінгвістичні поняття, зокрема, поняття тексту.

Під текстом (лат. *textus* – зв'язаність, матеріал, сплетення), письмовим або усним, прийнято розуміти логічно зв'язану послідовність лінгвістичних знаків. Основні характеристики тексту: зв'язаність; свідомість; цілісність (текст повинен бути закінчений за змістом). Крім того, класично текст має лінійну структуру.

Із середини 80-х років ХХ століття, коли швидкими темпами стали розвиватися комп'ютерні і телекомунікаційні мережі (WWW, Internet) і разом з тим глобальна комунікація, у науковий побут увійшло нове поняття гіпертексту (Hypertext). Виявилось, що для рішення багатьох інформаційних завдань нелінійний і мультимедійний характер тексту може бути ефективніше. Гіпертекст – це технологія організації інформації та особливим чином структурований текст, розбитий на окремі блоки, що має нелінійне подання, для ефективної презентації інформації в комп'ютерних середовищах. Для стандарту подання інформації в мережах стали розроблятися спеціалізовані мови. Наприклад, HTML – HyperText Markup Language – мова гіпертекстової розмітки документів.

При роботі з текстом можна вирішувати безліч завдань від загальнофілологічних до статистичних. Умовний напрямок обробки тексту за допомогою доступних комп'ютерних та інформаційних технологій можна назвати автоматичною обробкою тексту, тобто це будь-яке перетворення тексту на природній або штучній мовах за допомогою комп'ютера. Автоматична обробка текстів здійснюється, як ми помітили раніше, у технологіях інформаційного пошуку (при індексуванні текстів), машинного перекладу, словниковій роботі. У цьому розділі ми розглянемо деякі незгадані раніше додатки та завдання.

1. Автоматизовані операції по друку, редагуванню і верстці тексту.

Спочатку користувачеві були доступні примітивні операції по обробці тексту в простих комп'ютерних програмах-редакторах (Lexicon), згодом вимоги до редагування і подання документів зростали, що призвело до створення вдосконалених систем типу WordProcessors (Microsoft Word та ін.). Зараз це розвинені програми з функціями обробки таблиць, графіки, перевірки орфографії й стилістики. Крім того, є спеціальні видавничі системи для професійної верстки документів, газет, рекламної продукції, книг (наприклад, PageMaker або просунуті пакети типу Adobe® Digital Publishing Suite).

2. Розпізнавання тексту.

Розпізнавання тексту в цьому змісті прив'язано до завдання розпізнавання символів (Character Recognition) і реалізується ефективніше для друкованого тексту, чим для рукописного. З роботою лінгвістичного розпізнавача користувач зіштовхується при скануванні тексту, наприклад, через платформу ABBYY FineReader. У цей час завдання розпізнавання друкованого тексту практично вирішено у силу обмеженого набору друкованих гарнітур. Для рукописного тексту (технологію називають ICR, тобто Intelligent Character Recognition) розпізнавання графем набагато складніше, тому що потрібні більш доконалі алгоритми розпізнавання образів.

3. Автоматичне реферування та анотування текстів.

Автоматичне реферування (Automatic Text Summarization) – це складання коротких викладів матеріалів, анотацій або дайджестів, тобто витяг найбільш важливих відомостей з одного або декількох текстових документів і генерація короткого змісту аналізованого

об'єкта. Анотування, на відміну від реферування, припускає ще більший стиск змісту вихідного тексту. Існує багато способів рішення цього завдання прикладної лінгвістики. Як правило, автоматичне реферування засноване на вибірці ключових фрагментів документів – виділенні найбільш інформативних фраз та елементів – і їхній збірці. При автореферуванні активно використовується статистична інформація з інформативності різних елементів тексту по частоті їхньої появи в ньому, інформативність елемента тексту також залежить від його позиції в документі або текстових маркерах (наприклад, терміни виділені курсивом або напівжирним), які характеризують їхню змістовну значимість. Існують як комерційні, так і відкриті, наприклад, Open Text Summarizer, окремі продукти для реферування, а також вбудовані функції автореферату (наприклад, в MS Word).

#### 4. Корпусна лінгвістика.

Корпусна лінгвістика (Corpus Linguistics) – цілий напрямок прикладної лінгвістики, що активно розвивається з кінця XX століття, а в Україні – з початку XXI століття. У цей час електронних корпусів текстів для різних мов і додатків існує велика безліч.

Завдання корпусної лінгвістики пов'язані з розробкою технологій побудови електронних лінгвістичних ресурсів особливого типу – корпусів текстів (Corpora) – і рішенням завдань різного роду на базі цих текстів. В основному, такі колекції і масиви текстів відбивають реальне функціонування тієї чи іншої мови, а їхній перенос у комп'ютерні середовища активізував їх практичне і широке використання в прикладній лінгвістиці.

Корпусна лінгвістика дає матеріал для різного роду досліджень мови та його варіантів і визначає основний метод аналізу текстів і мови на базі корпусів. Однією з важливих особливостей методу аналізу на базі корпусів є дослідження не тільки чисто лінгвістичних явищ (граматичних або лексичних функцій слів, їхніх зв'язків з іншими лексемами), але і таких явищ, як, наприклад, частотності лексем або граматичних конструкцій у тих або інших жанрах, діалектах.

Корпусний підхід, або метод лінгвістичного дослідження, заснований на корпусах текстів, орієнтований на прикладне вивчення мови, його функціонування в реальних середовищах і

текстах, що важливо, наприклад, для викладання мови і для комп'ютерної лінгводидактики.

## *2. Сучасна лінгводидактика.*

У першій лекції ми вже звертали увагу на те, що англомовний термін «Applied Linguistics» за кордоном часто трактується трохи вужче, ніж ми звикли в Україні, і під ним розуміється напрямок, пов'язаний з методами і технологіями викладання мови, тобто лінгводидактика. Цей погляд на сферу інтересів прикладної лінгвістики існує і у нас в країні.

Одним із сучасних напрямків прикладної лінгвістики є можливість автоматизованого навчання іноземним мовам і підтримки його викладання. За кордоном цей напрямок, відомий як Computer-Assisted Language Learning and Teaching (CALL/CALT), є перспективним у силу багатьох об'єктивних причин і викладається як спеціальна дисципліна прикладного мовознавства на педагогічних і лінгвістичних факультетах коледжів та університетів.

Застосування комп'ютерних мультимедійних, дистанційних і відкритих технологій у процесі навчання іноземним мовам позитивно зарекомендувало себе в останні десятиліття. Мультимедіа технології дозволяють моделювати середовище, що імітує лінгвістичну і комунікативну реальність, що дуже важливо для мовного навчання, а також активізувати основні методичні принципи навчання іноземним мовам – розвиток навичок: аудіювання, говоріння, читання і листування. Невід'ємною частиною комп'ютерного утворення із впровадженням нових інформаційних і мережних технологій стали такі електронні засоби його підтримки, як машинні перекладачі, словники, національні корпуси текстів. Величезний потенціал для CALL/CALT несе доступ в Internet, у якому можна знайти мільйони різних ресурсів для підвищення ефективності іншомовної освіти, а також різні технології дистанційної комунікації, наприклад, Skype.

### Запитання для самоперевірки

1. Дати загальну характеристику понять тексту та гіпертексту.



2. Охарактеризувати загальні види автоматичної обробки тексту.

3. Визначити роль і місцлінгводидактики та CALL/CALT.  
Література: [2-3, 9, 17-18, 20, 22, 27, 32, 36].

## 2. Основні терміни та поняття

FAMT (Fully-Automated Machine Translation) – повністю автоматичний машинний переклад.

HAMT (Human-Aided Machine Translation) – машинний переклад за участю людини.

HTML (HyperText Markup Language) – мова гіпертекстової розмітки документів.

MANТ (Machine-Aided Human Translation) – переклад, здійснюваний людиною із залученням допоміжних програмних і лінгвістичних засобів.

Гіпертекст – це технологія організації інформації та особливим чином структурований текст, розбитий на окремі блоки, що має нелінійне подання, для ефективної презентації інформації в комп'ютерних середовищах.

Граматики – наука про способи і засоби побудови та зміни слова, про способи і засоби побудови речення, підрозділяється на дві області – морфологію і синтаксис.

Дескрипторна мова – найбільш природна і популярна форма мов індексування для вираження пошукових образів і у теперішній час широко використовується в сучасних інформаційно-пошукових системах (зокрема в багатьох пошукових системах мережі Internet).

Дискурс – наука, що вивчає та аналізує характерні для даного виду текстів (або комунікантів) особливості лексики, синтаксису, семантики або прагматики мовних одиниць, що проявляються в актуальних комунікативних актах, мовленні і письмових текстах (наприклад, спортивний дискурс).

Знак-індекс – вказівний знак, у якому означуване та означаюче зв'язані між собою причинно-наслідковими відносинами.

Знак-символ – знак, у якому означуване та означаюче зв'язані між собою в рамках деякої конвенції, тобто за попередньою домовленістю.

Кіничний знак – це знак, що володіє рядом властивостей, властивих позначуваному їм об'єкту

Інформатика (Computer Science) – наука про закономірності запису, зберігання, переробки, передачі та використання інформації за допомогою технічних засобів.

Інформаційно-пошукова система (ІПС) – це упорядкована сукупність документів (масивів документів) і інформаційних технологій, призначених для зберігання і пошуку інформації – текстів (документів) або даних (фактів). Інформаційно-пошуковими системами є будь-які певним чином організовані сховища інформації. Причому інформаційно-пошукові системи можуть бути і неавтоматизованими. Головне – це цільова функція: зберігання і пошук інформації.

Лексикологія – наука, що вивчає словниковий склад мови і закономірності його розвитку.

Лінгвістика (мовознавство) – це наука про природну людську мову взагалі і про всі мови світу як індивідуальних його представників

Машинний переклад – це перетворення комп'ютером тексту на одній природній мові в еквівалентний за змістом текст на іншій природній мові.

Мова програмування – це клас штучних мов, призначених для обробки інформації за допомогою комп'ютера. Будь-яка мова програмування – це строга (формальна) знакова система, за допомогою якої записуються комп'ютерні програми.

Морфологія – навчання про способи і засоби побудови та зміни слів.

Образотворчий знак – це знак, у якому значення має природно йому властивий вираз.

Повнотою пошуку (Recall) називається міра, що обчислюється як відношення кількості виданих релевантних документів до загального числа релевантних документів, що містяться в базі інформаційно-пошукової системи.

Прагматика – навчання про умови і цілі комунікації, що впливають на розуміння, так називане «вивчення мови в контексті», вивчення відносин між засобами мови і тими, хто цими засобами користується.

Програма – це структурно строгий текст, записаний за

формально заданими правилами штучної мови програмування.

Семантика – наука про зміст (значеннях) слів і речень, частин мови та членів речень.

Семіотика (греч. semeon – знак) – це наука про загальну теорію знаку, що досліджує будь-які знакові системи як засоби позначення і передачі значення або інформації.

Синтаксис – навчання про способи і засоби побудови речень.

Системи машинного перекладу (МП) – це цілий комплекс спеціальних найскладніших програм та алгоритмів плюс спеціальні автоматичні словники і формалізовані граматики для кожної мовної пари (вхідної і вихідної мов).

Словник – певним чином організована збірка слів, звичайно із приписаними їм коментарями, в яких описуються особливості їхньої структури та/або функціонування.

Тезаурус – це особливого роду словник, організований, по-перше, за тематичним принципом, і, по-друге, за принципом ієрархії відносин і «від змісту до слова», тобто ідеографічний словник орієнтований на семантику мови, і кожен такий словник – це деяка семантична модель лексики, побудована на ієрархічних відносинах типу «рід-вид», «частина-ціле», «синоніми» і т.ін.

Текст (лат. textus – зв'язаність, матеріал, сплетення) письмовий або усний – прийнято розуміти логічно зв'язану послідовність лінгвістичних знаків. Основні характеристики тексту: зв'язаність; свідомість; цілісність (текст повинен бути закінчений за змістом).

Точність пошуку (Precision) – це відношення кількості виданих системою релевантних документів до загального числа документів у видачі.

Умовний знак – це знак, у якому зв'язок між виразом та змістом встановлено за згодою групи людей, «внутрішньо не мотивований».

Фонетика і фонологія – наукові напрямки мовознавства, що вивчають звукову систему мови та особливості її функціонування.

Формальна грамика – це система строгих (часто математичних) правил, що дозволяє за допомогою однакових процедур одержувати (виводити) правильні вирази даної мови або аналізувати наявні вирази на предмет їхньої відповідності правилам мови.

### 3. Рекомендовані джерела інформації

#### Основна література

1. Батура Т.В. Математическая лингвистика и автоматическая обработка текстов на естественном языке : учебное пособие. – Новосибирск: РИЦ НГУ, 2016. – 166 с.

2. Большакова Е.И. и др. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : Учеб. пособие / Авторы: Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. М.: МИЭМ, 2011. – 272 с.

3. Дарчук Н.П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту) : Підручник. – К.: Видавничо-поліграфічний центр "Київський університет", 2008. – 351 с.

4. Захаров В.П., Богданова С.Ю. Корпусная лингвистика : учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с.

5. Маннинг К., Рагван П., Шютце Х. Введение в информационный поиск. – М.: 2011.

6. Николаев И.С., Митренина О.В., Ландо Т.М. Прикладная и компьютерная лингвистика. – М.: Ленанд, 2016. – 320 с.

7. Партико З.В. Прикладна і комп'ютерна лінгвістика. Вступ до спеціальності : навчальний посібник / З. В. Партико. – Львів: Афіша, 2008. – 224 с.

8. Рассел С., Норвиг П. Искусственный интеллект. Современный подход. – 2-е изд. – М.: Вильямс, 2007. – 1410 с.

9. Соснина Е.П. Введение в прикладную лингвистику : учебное пособие / Е.П. Соснина. – 2-е изд., испр. и доп. – Ульяновск : УлГТУ, 2012. – 110 с.

#### Допоміжна

10. Анисимов А.В. Компьютерная лингвистика для всех: Мифы. Алгоритмы. Язык. – К., 1991.

11. Баранов А.Н. Введение в прикладную лингвистику : Учебное пособие. – М.: Эдиториал УРСС, 2001. – 347 с.

12. Баранов А.Н. Категории искусственного интеллекта в лингвистической семантике. Фреймы и сценарии. – М., 1987.

13. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии. – М.: Русский мир, 2004.

14. Білевич Т.Л. Принципи створення машинної версії “Словника староукраїнської мови XIV-XV ст.” // Проблеми українізації комп’ютерів. – Львів, 1993. – С.14-15.

15. Брага І.І. Мовна репрезентація образу держави у пресі України (кінець 70-х – початок 2000-х років). – Автореф. дис. ... канд. філол. наук. – К., 2002.

16. Ван Дейк Т.А. Язык. Познание. Коммуникация. – Благовещенск: БГК им. И. А. Бодуэна де Куртенэ, 2000. – 308 с.

17. Вінцюк Т. Комп’ютерні автоматичні системи розпізнавання та синтезу українського мовлення // Проблеми українізації комп’ютерів. – К., 1993. – С.21-32.

18. Вінцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов. – К., 1987.

19. Волошин В.Г. Комп’ютерна лінгвістика. Навчальний посібник : Навч. посібник. – Суми: Університетська книга, 2004. – 382 с.

20. Гальскова Н.Д., Гез Н.И. Теория обучения иностранным языкам. Лингводидактика и методика : учебное пособие. – 3-е изд., стер. – М.: Академия, 2006. – 336 с.

21. Голубовская И.А. Этнические особенности языковых картин мира. – К., 2002. – 293 с.

22. Грязнухіна Т.О., Нікула М.В. Система автоматичного морфологічного аналізу українського наукового тексту // Проблеми українізації комп’ютерів. – К., 1993. – С.42-46.

23. Дарчук Н.П., Грязнухіна Т.О. Частотний словник сучасної української публіцистики // Мовознавство. – 1996. – № 4-5. – С.15-18.

24. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике. – М.: Академия, 2004. – 208 с.

25. Карасик В.И. Языковой круг: личность, концепты, дискурс // Монография. – Волгоград: Перемена, 2002. – 477 с.

26. Карпіловська Є.А. Вступ до прикладної лінгвістики: комп’ютерна лінгвістика : підручник / Є. А. Карпіловська. – Донецьк: ТОВ «Юго-Восток, Лтд», 2006. – 188 с.: іл.

27. Карпіловська Є.А. Машинні версії традиційних словників як основа для укладання комп’ютерних словників та тезаурусів // Мовознавство. – 1996. – № 4-5. – С.19-22.

28. Кобозева И.М. Лингвистическая семантика. – М.: Эдиториал УРСС, 2000. – 352 с.

29. Ковтуненко Л.С. Комп'ютерні аспекти лексикографічних систем // Мовознавство. – 1996. – № 4-5. – С.28-34.

30. Кубрякова Е.С. Язык и знание. На пути получения знаний о языке: части речи с когнитивной точки зрения. Роль языка в познании мира / Рос. академия наук. Ин-т языкознания. – М.: Языки славянской культуры, 2004. – 560 с.

31. Марчук Ю.Н. Компьютерная лингвистика. – М.: АСТ: Восток-Запад, 2007. – 317 с.

32. Пиотровский Р.Г. Лингвистический автомат и его речемыслительное обоснование. – Минск, 1999.

33. Синтаксический анализ научного текста на ЭВМ. – К., 2000.

34. Тер-Минасова С.Г. Язык и межкультурная коммуникация. – М., 2000.

35. Тузов В.А. Компьютерная семантика русского языка. – Изд-во СПбГУ, 2003. – 400 с.

36. Klymenko N.F., Karpilovs'ka E.A. Computer Morpheme-Word-Formative Database of the Ukrainian Language and Its Applications // Journal of Quantitative Linguistics. – 1994. – Vol.1. – No.2. – P.113-131.

### **Інформаційні ресурси**

1. Бібліотека Харківської державної академії культури (<http://lib-hdak.in.ua/>)
2. Харківська державна наукова бібліотека імені В. Г. Короленка (<http://korolenko.kharkov.com/>)
3. Національна бібліотека України ім. В.І. Вернадського (<http://www.nbuv.gov.ua/>)
4. Національна бібліотека України ім. Ярослава Мудрого (<http://www.nplu.org/>)

## **ПРИКЛАДНА ЛІНГВІСТИКА**

Робоча програма навчальної дисципліни до курсу  
зі спеціальності 029 «Інформаційна, бібліотечна та архівна справа»

Укладач:

кандидат технічних наук, доцент ***Брусенцев В.О.***

*Друкується в авторській редакції*

План 2018

Підписано до друку 10.12.2018. Формат 60x84/16.

Гарнітура «Times». Папір для мн. ап. Друк ризограф.

Ум. друк. арк. 2,10. Обл.-вид. арк.2,20. тираж 100. Зам. №

ХДАК, 61057, Харків-3, Бурсацький узвіз, 4.